Máster Universitario en Investigación e Innovación en TIC
Universidad Autónoma de Madrid
Escuela Politécnica Superior
Departamento de Ingeniería Informática

Estudio de Técnicas de análisis y clasificación de señales EEG en el contexto de Sistemas BCI *(Brain Computer Interface)*

Memoria de Trabajo de Fin de Máster
presentada por

Claudia Nureibis Henríquez Muñoz

Dirigida por

Francisco de Borja Rodríguez Ortiz
Prof. Titular de Ciencias de la Computación e Inteligencia Artificial

Marzo 2014

"La vida, la naturaleza, la humanidad sólo son bellas palabras cuando son transformadas por un cerebro creador."

Edmond Jaloux

Agradecimientos

A Dios por ser mi compañía y guía a lo largo de mi vida, por ser mi fortaleza en los momentos de debilidad, y por brindarme una vida llena de aprendizajes, experiencias y sobre todo felicidad. Por permitirme realizar mis sueños.

A mi tutor Francisco, por su infinita paciencia, por haber estado siempre dispuesto a aclarar mis dudas y ayudarme, y por todo lo que he aprendido de él.

A mi esposo Mauricio, por todo, por estar ahí, por su apoyo constante y amor incondicional.

A mis padres Hermes y Zita, por su apoyo y cariño incondicional siempre, por haberme enseñado la virtud del esfuerzo y haber hecho de mi la persona que soy. A mis dos hermanos Leidy y Sergio por ser parte importante de mi vida. Por su apoyo, confianza y ánimo.

Y por último agradecer a los profesores y compañeros del Máster Universitario en Investigación e Innovación en Tecnologías de la Información y las Comunicaciones y en especial a mi amigo Iván por su apoyo y amistad durante todo el año.

Resumen

Las Interfaces Cerebro Computador (BCI) son una tecnología basada en la adquisición y procesamiento de señales cerebrales para el control de diversos dispositivos. Su objetivo principal es proporcionar un nuevo canal de salida al cerebro del usuario que requiere un control adaptativo voluntario. Usualmente los BCI se enfocan en reconocer eventos que son adquiridos por métodos como el Electroencefalograma (EEG). Dichos eventos representan distintos estados mentales, que deberán identificarse como distintas clases mediante un sistema de clasificación. En BCI, tras la adquisición de las señales cerebrales se procede a prepararlas para su posterior procesamiento. Para extraer y seleccionar las características se aplican diversos métodos, en los que, en función del tipo de señales de control que se utilice, será adecuado identificar un subconjunto que optimice las tareas para su clasificación. La selección de las características más discriminativas es imprescindible a la hora de diseñar BCI funcionales.

En este trabajo hemos realizado un estudio del estado del arte de BCI, en concreto, de los métodos más relevantes de procesado de la señal EEG utilizados en la etapa de extracción y clasificación de características. El uso y diseño de dichos métodos, es llevado a cabo de forma experimental mediante un análisis *offline* y un análisis *online simulado*, con el propósito de asegurar un mayor éxito en el desempeño del sistema BCI en tiempo real. Ambos tipos de análisis son necesarios y complementarios para el diseño de algoritmos efectivos, en el que el usuario recibe *feedback* del BCI. Usando la técnica de análisis que hemos propuesto *online simulado*, con la que los resultados de las pruebas realizadas son emulados para obtener una estimación de los mismos como si el usuario hubiera utilizado el sistema en tiempo real, fue posible minimizar el tiempo de detección de eventos y así optimizar el tiempo de respuesta del BCI.

Para el desarrollo práctico de este estudio, utilizamos una colección de datos reales de EEG relacionados con los Potenciales Corticales Lentos (SCP),

que contiene dos experimentos diferentes denominados "*Ia*", provenientes de sujetos sanos y "*Ib*" provenientes de sujetos con la enfermedad ALS. Dichos datos, se obtuvieron de la competición "BCI Competitions II" de 2003 [2].

Se han analizado diferentes metodologías para extraer las características más relevantes de estas señales SCP. En una primera fase, hemos realizado la clasificación con cada una de las técnicas estadísticas y métodos de extracción de características por separado para reducir la dimensionalidad. En una segunda fase las técnicas y métodos con los que obtuvimos mejores resultados, se han usado para conformar nuevos patrones (súper vectores de características) que se han utilizado en los diferentes métodos de clasificación estudiados. Con ello, minimizamos las tasas de error obtenidas, mejorando el rendimiento en la clasificación y superando a los resultados obtenidos por los ganadores de "BCI Competitions II" tanto en "*Ia*" [73] como en "*Ib*" [26].

Palabras claves: BCI, EEG, señales de control, SCP, extracción y clasificación de características, análisis *offline*, análisis *online simulado*, generación de súper vectores de características y tiempo real.

Abstract

BCI is a new technology based in the acquisition and signal processing to control a variety of devices. Its main goal is to provide a new output channel to the user's brain which requires a voluntary adaptive control. Usually, BCI is focused in recognizing acquired events by methods such as EEG. Those events represent different mental states which should be identified as different classes through a classification system. In BCI after the Brain signal acquisition it is necessary to process the signal, before processing it. To extract and select features it is necessary to apply a different method, in which, it will be suitable to identify a subset according to the signals control used to optimize the classification of tasks. To design functional BCI, it is indispensable to select the most discriminative features.

In this work we have made a BCI state of the art research, especially about the most relevant methods to process the EEG signal, used in the extraction and classification of characteristics steps. The use and design of those methods, has been made through an offline and simulated online analysis, in order to ensure a better performance of the BCI system in real time. Both kind of analysis are complemented and needed to design effective algorithms where the user receives feedback from the BCI. Using the analysis technique that we proposed, the simulated online analysis, where the results of the test are emulated to obtain an estimation as if the user had used the system in real time, it was possible to minimize the events detection time, and in consequence, optimizing the BCI response time.

To develop this study, we have used a collection of EEG real data related with Slow Cortical Potential (SCP), that contains two different experiments called "*Ia*" from healthy people, and "*Ib*" from people who suffer from ALS. Those data have been obtained from the "BCI Competitions II" held in 2003 [2].

We have analyzed different methodologies to extract the most relevant characteristics of those SCP signals. In a first phase, we made a classification with each one of the statistical techniques and extraction methods separately to reduce dimensionality. In a second phase, the techniques and methods, which we obtained the best results were used to make up new patterns (the super vectors of characteristics), that have been used in the different classification methods studied. Using those phases, we minimized the obtained error rates, improving the performance in the classification and improving the results obtained for the winners of "BCI competitions II" in both "*Ia*" [73] and "*Ib*" [26].

Key words: BCI, EEG, control signals, SCP, features extraction and classification, offline analysis, simulated online analysis, super vectors and real time.

Índice general

Índice de figuras

Índice de tablas

Parte I

Introducción

Capítulo 1

Introducción

1.1. Ámbito del Trabajo Fin de Máster

El interés por el desarrollo de las Interfaces Cerebro Computador se remonta a varias décadas atrás y responde a la necesidad del mundo moderno brindar una nueva forma de comunicación entre las personas y el computador, conllevando a una accesibilidad más amplia de uso del mundo digital y a una mayor automatización las tareas cotidianas humanas.

Los sistemas BCI son un campo en evolución continuo y creciente. En las últimas décadas el interés entre los grupos de investigación de todo el mundo ha ido aumentando. Cada año se acrecienta la participación tanto en congresos internacionales como en las competiciones BCI [115], [63].

Los grupos de investigación trabajan en el proceso de diseñar BCI óptimas que proporcionen una vía de comunicación a aquellas personas tetrapléjicas o "locked-in" (personas que han perdido el control muscular total o parcial). Jan Schüermann, una mujer de 53 años que fue diagnosticada con una enfermedad cerebral degenerativa hace 13 años y está paralizada del cuello para abajo, ha sido capaz de operar un brazo robótico con un nivel de control y fluidez nunca antes visto en este tipo de prótesis avanzadas. Expertos califican el suceso como un notable avance para las prótesis controladas directamente con la actividad cerebral [118]. Otros sistemas ya han permitido a pacientes paralíticos mecanografiar o escribir a mano, simplemente con imaginar las letras que quieren plasmar. Hasta hace poco, investigadores en EE. UU. usaron electrodos implantados directamente en la retina para permitir leer a un paciente invidente [7]. Las últimas tendencias en BCI, muestran a Zac Vawter, el primer hombre en controlar una extremidad artificial que usa

neuroseñales de los músculos superiores de la pierna que controlan la rodilla protésica y el tobillo [6]. Así mismo, Dennis Sorensen pasará a la historia por ser la primera persona con una mano protésica con un cable conectado a los nervios de su brazo, que ha vuelto a tener información sensorial en tiempo real [8]. Por último mencionar, que por primera vez en la historia un joven con parálisis, usará un esqueleto controlado por un BCI para patear el balón en la ceremonia de apertura de la Copa Mundial de fútbol (Brasil 2014) [14]. El uso de BCI pretende mejorar la calidad de vida de este tipo de personas.

Existe además, un gran interés por el desarrollo de nuevas tecnologías de control y comunicación por parte de la industria del entretenimiento, casas inteligentes y aplicaciones militares, lo que ha impulsado un incremento en la inversión para realizar este tipo de trabajos de investigación, sobrepasando el objetivo principal de un BCI [111], [74], [61], [18].

Los sistemas BCI utilizan diferentes técnicas de neuroimagen para registrar la actividad cerebral y traducir determinadas características, correspondientes a las intenciones del usuario, en comandos de control de un dispositivo [115]. El EEG es la modalidad más empleada para registrar las señales cerebrales, debido a que es una técnica no invasiva, fácil de manejar, de bajo coste y portable [116]. De esta actividad cerebral, se ha logrado extraer varias señales de control con el fin de determinar las intenciones del usuario. Estas señales son los Potenciales Evocados Visuales (VEP) [112], [43], [83], Potenciales Evocados P300 [36], [17], Potenciales Corticales Lentos (SCP) [81], [22] y Ritmos Sensoriomotores (μ y β) [72], [88].

Una vez que las señales registradas se muestrean y digitalizan se procede a preparar la señal para su procesamiento. Por medio de la aplicación de distintos métodos de extracción se pueden obtener determinadas características de la señal EEG que pueden ser útiles para discriminar entre diferentes tareas mentales. Después, mediante el empleo de métodos de selección se escoge el subgrupo de características más relevantes. Finalmente, los métodos de clasificación permiten determinar, a partir de las características seleccionadas, a qué clase de tarea mental pertenece una muestra concreta.

En este Trabajo de Fin de Máster (TFM) se estudiaran los métodos de extracción y clasificación de características de las señales EEG más utilizados actualmente en BCI. Para el desarrollo del TFM se usará el conjunto de señales reales de EEG propuesto por "BCI Competitions II" en el año 2003 [2] que contiene dos diferentes experimentos: "Ia" y "Ib". Se estudiará específicamente las señales de EEG relacionadas con la autorregulación cortical.

Este tipo de señales son conocidas como Potenciales Corticales Lentos. Finalmente, se mostrarán los resultados obtenidos en la clasificación, luego de realizar el procesado de los datos SCP en los escenarios de análisis *offline* y *online simulado*, obteniendo un porcentaje de error menor y mejor que los resultados obtenidos por los ganadores de "BCI Competitions II" [2] para "*Ia*" [73] y "*Ib*" [26].

1.2. Objetivos

El objetivo principal de este TFM es realizar un estudio teórico y práctico de los métodos más relevantes en las distintas etapas de procesamiento para las señales de control SCP en los sistemas BCI en el escenario *offline* y el estudio, análisis y discusión de cómo se podrían extrapolar dichos métodos a un escenario *online simulado*. Para abordar este problema, se esperan lograr cuatro hitos fundamentales:

1. Estudiar el estado del arte de los sistemas BCI.

2. Estudiar e identificar los métodos más relevantes de procesado de la señal EEG utilizados en la etapa de extracción y clasificación de características.

3. Aplicar y evaluar prácticamente los diferentes métodos de extracción y clasificación de características. Para ello, se utilizará una colección de datos reales de EEG relacionados con los Potenciales Corticales Lentos (SCP), que contiene dos experimentos diferentes denominados *"Ia"* y *"Ib"* proporcionados por "BCI Competitions II" de 2003.

4. Diseñar y aplicar metodologías propias en el procesado de las señales SCP, que contribuyan en el éxito del desempeño en la clasificación, como son, (i) la generación de súper vectores apropiados para la clasificación de eventos para comandar un sistema BCI y (ii) el desarrollo de la técnica de análisis *online simulado* que optimiza el tiempo de respuesta del BCI.

1.3. Estructura del Trabajo Fin de Máster

Esta memoria se divide en seis capítulos incluyendo este primer capítulo introductorio.

El segundo capítulo aborda conceptos fundamentales y necesarios para adentrarse en el campo de los sistemas BCI y además hace un estudio de las investigaciones más actualizadas. Seguidamente, se explican las diferentes técnicas de neuroimagen profundizando en EEG. También, se estudian las diferentes técnicas utilizadas para el procesamiento de señales cerebrales. Estas técnicas engloban aquellas que pretenden extraer la información característica de las señales cerebrales y las que emplean esta información extraída para clasificar las señales con el fin de conocer las intenciones del usuario. Finalmente, se motiva a la investigación en BCI mencionando los retos y el futuro del mismo.

El tercer capítulo describe en detalle los conjuntos de datos *"Ia"* y *"Ib"* de "BCI Competitions II" [2] relacionados con la autorregulación de SCP, que son los empleados para realizar el estudio de métodos de extracción y clasificación de características. Se resume también, los métodos para el procesamiento de las señales aplicados por los ganadores de "BCI Competitions II" en dichos conjuntos de datos. Por último, se busca extraer las características de las señales SCP, mediante un primer análisis visual y estadístico de las señales.

El cuarto capítulo presenta los métodos de extracción y clasificación de características empleados en los conjuntos de datos "*Ia*" y "*Ib*" de SCP en el escenario *offline* y su extrapolación en el escenario *online simulado*. Los métodos que se usaron para extraer características de las señales de autorregulación cortical son: Análisis de Componentes Principales (PCA), modelos autorregresivos (AR) y Transformada Wavelet Discreta (DWT). Para la clasificación de las características extraídas se utiliza Análisis Discriminante Lineal (LDA) y una Máquina de Vectores de Soporte (SVM).

El quinto capítulo incluye la discusión de los resultados obtenidos una vez que se han analizado las señales de SCP de "BCI Competitions II" [2] y se han realizado todas las pruebas con las técnicas y métodos explicados en el tercero y cuarto capítulo.

Finalmente, el sexto capítulo contiene las principales conclusiones extraídas de los resultados obtenidos uniéndolos con los objetivos inicialmente propuestos y las líneas de investigación que puedan surgir del mismo.

Parte II

Interfaces Cerebro Computador

Capítulo 2

Fundamentos de los sistemas BCI

En este capítulo se presentan los fundamentos considerados esenciales para el seguimiento de este TFM. Inicialmente se estudian conceptos específicos y necesarios para adentrarse en las Interfaces Cerebro Computador. Seguidamente, se aborda las señales de entrada a los sistemas BCI, más concretamente, las señales de EEG aunque se presentarán otras señales electrofisiológicas consideradas relevantes. Finalmente se detallarán otros aspectos importantes de los sistemas BCI y se revisa los métodos en tiempo, frecuencia y tiempo-frecuencia más populares y utilizados actualmente para extraer las características relevantes de las señales cerebrales y su posterior clasificación.

2.1. ¿Qué es BCI?

Las Interfaces Cerebro Computador (*Brain Computer Interface*, BCI) es un medio de comunicación basado en la actividad neuronal generada por el cerebro y es independiente de sus vías de salida a nervios periféricos y músculos. El objetivo principal de los sistemas BCI no es determinar la intención de una persona visualizando la actividad cerebral, sino que es el de proporcionar un nuevo canal de salida al cerebro del usuario que requiere un control adaptativo voluntario [114].

El caso más importante de las aplicaciones BCI es en el contexto clínico, desarrollado para pacientes con Esclerosis Lateral Amiotrófica (*Amyotrophic Lateral Sclerosis*, ALS), Lesión de la Médula Espinal (*Spinal Cord Injury*, SCI) y otros trastornos de la función del movimiento, con el fin de controlar dispositivos, tales como una silla de ruedas o un robot. En estos, la entrada es

una señal cerebral obtenida a partir de EEG o alguna otra medición directa de la actividad cerebral, y la salida es tomada como una señal de control [79]. Más allá de este noble fin, los sistemas BCI ofrecen además nuevas técnicas para el entretenimiento, casas inteligentes y aplicaciones militares [111], [77], [66].

2.2. Historia de la Investigación BCI

El cerebro siempre ha fascinado a los humanos, y en particular a un científico alemán llamado Hans Berger, quien descubrió la Electroencefalografía en humanos, y especulaba en su primera revisión de sus experimentos con el "Elektrenkephalogramm" en 1929, acerca de la posibilidad de leer pensamientos a partir de trazos de EEG usando sofisticados análisis matemáticos. Grey Walter, uno de los pioneros de EEG, robótica, cibernética e inteligencia artificial, construyó el primer analizador de frecuencia automático con la intención de discriminar los pensamientos y el lenguaje en el EEG humano [23].

En 1969, se hizo la primera publicación acerca de los estudios de condicionamiento operante de Fetz y sus colegas [39], en los que mostraron por primera vez, que los monos podían aprender a controlar voluntariamente las tasas de disparo de las neuronas individuales y múltiples en la corteza motora primaria, si eran recompensados por la generación de patrones apropiados de la actividad neuronal.

Sin embargo, solo el reciente desarrollo de los sistemas BCI, nos ha colocado más cerca al sueño de estos pioneros de la investigación en EEG.

Los sistemas BCI invasivos y no invasivos se originan desde diferentes investigaciones, aunque ambos tienen sus raíces en experimentos con animales. Los sistemas BCI invasivos constan de redes de micro-electrodos implantados en la corteza motora de pacientes paralizados [37] o en la corteza pre-motora de monos [29]. Este tipo de BCI, trata de reconstruir movimientos intencionales aprendidos a partir de patrones *online*. Nicolelis y su grupo [29] demostraron en monos después de un largo entrenamiento de movimientos de alcanzar y agarrar, que la ejecución de patrones de 32 neuronas son suficientes para ejecutar el movimiento directamente con un miembro artificial.

El interés en los sistemas BCI ha ido aumentado conforme han ido pasando los años. En 1995 había seis grupos de investigación dedicados al estudio de estos sistemas. En el primer congreso internacional, que tuvo lugar en el

año 2000, participaron 20 grupos de investigación [115]. En el tercer congreso el número de participantes aumentó hasta alcanzar casi el triple de los que participaron en el primer congreso [63]. Así mismo, en la Figura 2.1 se puede comprobar el espectacular aumento en la producción científica en BCI.

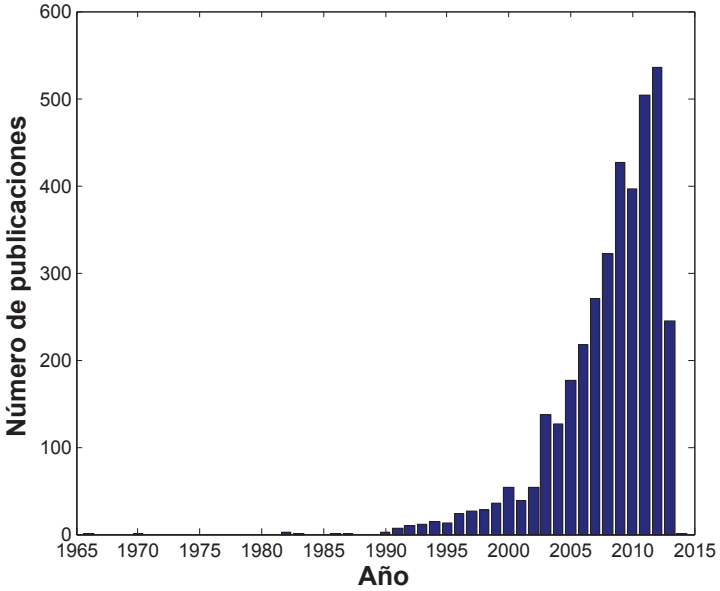

Figura 2.1: *Número de artículos en revista indexados en web of science*

2.3. Componentes funcionales de BCI

Un BCI detecta y mide las características de las señales cerebrales que revelan las intensiones del usuario y traduce estas características en tiempo real en comandos que logran la intención del usuario [15].

Existen diversas configuraciones con respecto a la implementación de un sistema BCI, sin embargo, Mason [71] propone un modelo general aplicable a cualquiera de estos sistemas, siendo dicho modelo el que se muestra en la Figura 2.2.

La función de cada uno de los componentes que forman dicho modelo se detalla a continuación:

- Sujeto: Persona que controla el dispositivo en el sistema BCI, modificando para ello intencionadamente su actividad cerebral con el fin de generar las señales de control que actúan sobre el dispositivo.

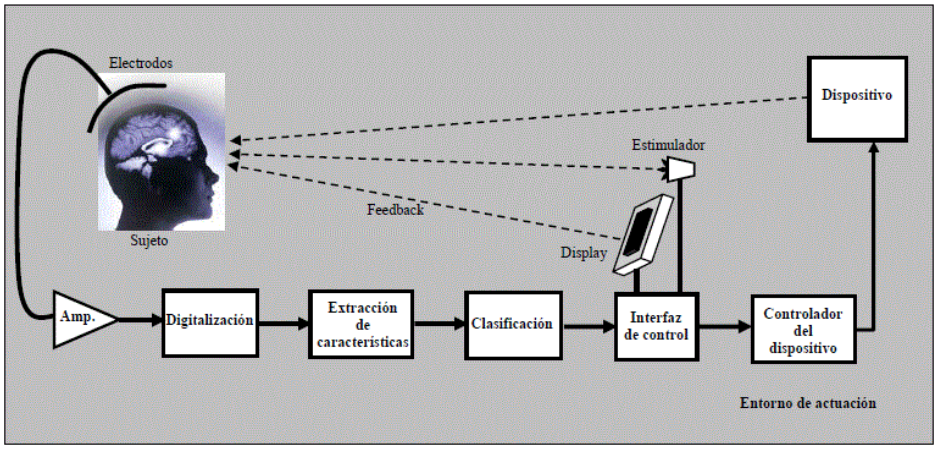

Figura 2.2: *Estructura general de un sistema BCI. Fuente: [71].*

- Electrodos: Convierten dicha actividad cerebral en señales eléctricas.
- Amplificador: Amplifica las señales eléctricas.
- Digitalización: Conversión de las señales analógico-digital, necesarias para el posterior procesado.
- Extracción de características: Se transforman las señales eléctricas ya amplificadas en parámetros que caractericen la actividad cerebral modificada por el sujeto.
- Clasificación: Se asignan a los parámetros que caracterizan las señales eléctricas, unas señales lógicas de control. En definitiva se trata de clasificar los diferentes parámetros de entrada entre un número determinado de estados posibles.
- Interfaz de control: Traduce las señales lógicas de control en señales de control apropiadas para un determinado dispositivo. La interfaz de control puede ir acompañada de un "display de control" que proporciona *feedback* (visual, auditivo o táctil) al sujeto sobre el resultado de su estado mental. Dicha información puede ser usada por el sujeto para modificar su actividad cerebral en caso de que la respuesta recibida no sea la adecuada.
- Controlador del dispositivo: Transforma las señales de control procedentes de la interfaz de control en las señales físicas necesarias para actuar sobre el dispositivo.

- Dispositivo: Como en cualquier sistema de control, el número de dispositivos sobre el cual puede actuar un sistema BCI puede ser ilimitado. Teniendo en cuenta que las distintas señales lógicas de control proporcionadas por estos sistemas son aún bastante limitadas, esto lleva a que se reduzca considerablemente el número de aplicaciones finales. El estado del dispositivo es a su vez una realimentación hacia el sujeto, proporcionándole información sobre el control que está ejerciendo sobre éste.

- Entorno de actuación: Hace referencia a las condiciones físicas del entorno sobre el cual se está usando el sistema BCI, así como los objetos y personas presentes en dicho entorno. Estos factores por lo general pueden ser controlados en el laboratorio.

En la literatura existen otros modelos de sistemas BCI, sin embargo, todos ellos son versiones simplificadas del modelo presentado. En éstos, al conjunto formado por los electrodos, el amplificador y la digitalización, se le suele agrupar en una sola etapa conocida como etapa de adquisición de la señal. Una nueva etapa conocida como pre-procesamiento, prepara las señales adquiridas en la etapa anterior en una forma adecuada para su posterior procesamiento. A las etapas de extracción y de clasificación se las suele incluir en un único bloque denominado procesamiento de la señal. Por último, la etapa de control del dispositivo externo, la forman la interfaz y el "display" de control.

Más adelante en la Sección 2.6, se hablará en detalle de cada uno de los bloques o etapas consecutivas de los sistemas BCI.

2.4. Enfoques en BCI: Señales para analizar

Un BCI puede monitorear variedad de señales del cerebro, como eléctricas, magnéticas y metabólicas. Se pueden medir las corrientes eléctricas, producidas por la sincronización de corrientes sinápticas, mediante Electroencefalografía (EEG), Electrocorticografía (ECoG) y mediante electrodos epidurales e intracorticales. Los campos magnéticos pueden ser grabados con Magnetoencefalografía (MEG), mientras que la actividad metabólica del cerebro que se refleja en los cambios en el flujo sanguíneo, puede ser observada con Tomografía por Emisión de Positrones (PET), Imágenes de Resonancia Magnética funcional (fMRI) y mediante técnicas ópticas como Espectroscopia de Infrarrojo Cercano (NIRS) [41].

Como se ha mencionado, la actividad cerebral puede ser registrada mediante modalidades invasivas y no invasivas. Las modalidades invasivas como ECoG o la adquisición de señales eléctricas en neuronas individuales [84] requieren la implantación quirúrgica de un "*array*" de electrodos con estabilidad a largo plazo, que pueden situarse en diferentes partes de la corteza cerebral para registrar la actividad de una neurona o pequeños grupos de estas. En el ámbito humano, las modalidades invasivas han sido aplicadas para movimientos de cursor y aplicaciones de escritura [109]. Tanto EcoG como la adquisición de señales eléctricas en neuronas individuales tienen mejor resolución que las modalidades no invasivas y presentan un menor nivel de ruido cuanto más invasivo sea el método.

PET es considerada como una modalidad parcialmente invasiva, ya que requiere una inyección radiactiva, aunque no se necesita cirugía [90].

Las modalidades no invasivas como la Tomografía Computarizada (*Computerized Tomography*, CT), Tomografía por Emisión de Fotón Único (*Single Photon Emission Computed Tomography*, SPECT), Imagen de Resonancia Magnética (*Magnetic Resonance Imaging*, MRI), fMRI, MEG y EEG [15], [98], no necesitan implantación quirúrgica y, por ello, las señales obtenidas tienen una Relación Señal-Ruido (*Signal to Noise Ratio*, SNR) menor. Aunque los BCI basados en este tipo de métodos tienen una velocidad y precisión modesta, estudios actuales se centran en el uso de sistemas multicanal (típicamente 64 canales) para poder mejorarlas [116].

Las técnicas de neuroimagen permiten el estudio del cerebro humano en vivo [48]. A continuación, se describen las técnicas de neuroimagen más usadas en BCI.

2.4.1. Electroencefalografía

La Electroencefalografía (*Electroencephalography*, EEG) mide la actividad eléctrica del cerebro usando electrodos adheridos al cuero cabelludo; que son activados por el flujo de corrientes eléctricas durante excitaciones sinápticas de las dendritas en las neuronas. Su rango se expresa en microVoltios debido a la alta atenuación por el cráneo y el cuero cabelludo (de tal forma que para un análisis adecuado las ondas de EEG tienen que ser amplificadas) [67]. Un ejemplo representativo de EEG para una configuración BCI se muestra en la Figura 2.3.

Figura 2.3: *Representación esquemática de un sistema de BCI basado en EEG. Usando el paradigma conocido como imaginación motora, sujetos humanos controlan el movimiento en 3D de un helicóptero virtual. Fuente: [4].*

Los diseños de BCI basados en EEG son muy útiles para dispositivos de control de manos libres y la comunicación, dado que estos usan la actividad eléctrica del cerebro para interactuar con el ambiente externo, sin involucrar el uso de músculos periféricos y extremidades. Algunas aplicaciones actuales de BCI en sistemas de comunicación son para que individuos con parálisis se comuniquen con su alrededor a través de la selección de un menú de caracteres y un dispositivo de control tales como el movimiento de una silla de ruedas, control de prótesis y tecnologías de asistencia [89].

Los sistemas BCI basados en EEG utilizan una estrategia de colocación de los electrodos sugerido por el Sistema Internacional 10/20 [60] que garantiza el cubrimiento de todas las partes de la cabeza, como se muestra en la Figura 2.4. Para una mejor resolución espacial, es también común el uso de una variante de este sistema que rellena los espacios entre los electrodos con electrodos adicionales [69].

Las posiciones exactas para cada electrodo son la intersección de las líneas calculadas desde las mediciones en puntos de referencia en cráneos estándar. La letra de cada electrodo identifica el segmento del lóbulo particular (FP lóbulo prefrontal, F, lóbulo frontal; T lóbulo temporal, C, lóbulo central, P, lóbulo parietal; o, lóbulo occipital). El numero de la segunda letra identifica su localización hemisférica (z, denotando la línea cero, se refiere que un

electrodo está colocado a lo largo de la línea media del cerebro; números pares representan el hemisferio derecho; números impartes representan el hemisferio izquierdo. Los números están en orden ascendente con distancia en incremento desde la línea media) [71].

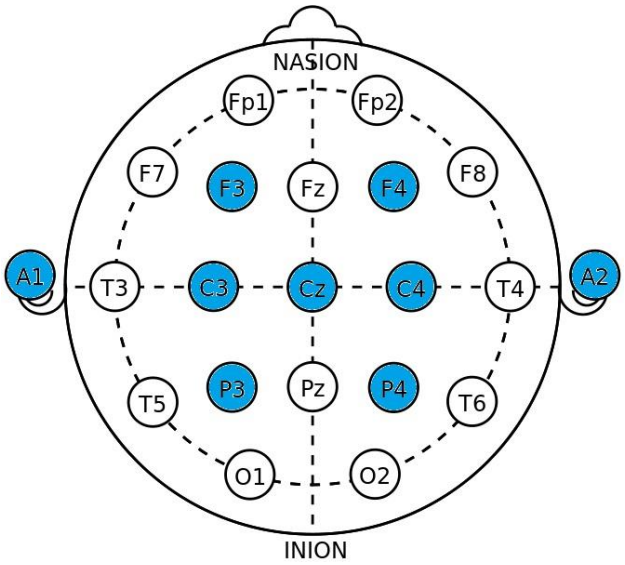

Figura 2.4: *Colocación de electrodos para la adquisición no invasiva de señales usando EEG. Fuente: [1].*

Los datos de SCP estudiados en este TFM, han sido registrados desde las posiciones que se encuentran resaltadas en azul en el mapa de electrodos de la Figura 2.4. Para el conjunto de datos *"Ia"*, se usaron 6 canales como sigue, canal 1: A1-Cz, canal 2: A2-Cz, canal 3: 2 cm frontal de C3, canal 4: 2 cm parietal de C3, canal 5: 2 cm frontal de C4 y canal 6: 2 cm parietal de C4. Para el conjunto de datos *"Ib"*, se utiliza los mismos 6 canales y un canal adicional para detectar los movimientos oculares verticales (vEOG).

La señal de EEG se mide como la diferencia de potencial a lo largo del tiempo entre la señal o electrodo activo y el electrodo de referencia. Un tercer electrodo adicional, conocido como el electrodo de tierra, se utiliza para medir la diferencia de voltaje entre los puntos activos y los puntos de referencia. La configuración mínima para la medición de EEG consta de de un electrodo activo, uno de referencia y un electrodo de tierra [84].

De la actividad cerebral detectada por los electrodos pueden realizarse dos tipos de medidas:

- Oscilaciones continuas de potencial, que en su conjunto forman lo que es EEG. A estas oscilaciones se les conoce como los ritmos de la actividad cerebral o ritmos cerebrales.

- Cambios de potencial que surgen como respuesta a la aparición de un determinado evento, el cual puede ser interno o externo al sujeto. A estos potenciales se les conoce como potenciales relacionados con eventos (ERP) y se derivan del propio EEG.

2.4.1.1. Ritmos cerebrales

Diferentes frecuencias eléctricas podrían estar vinculadas a las acciones y diferentes estados de la conciencia. Esto fue hecho observando a sujetos desarrollando diferentes tareas, como resolver problemas matemáticos, mientras se grababa su EEG.

La Figura 2.5 muestra las bandas de frecuencias más usadas a excepción de la banda MU (μ) (cuya frecuencia es igual que la banda alfa (α)) y sus relaciones de la actividad de las ondas del cerebro humano. Los rangos de frecuencias conocidas han sido definidos de acuerdo a la distribución sobre el cuero cabelludo o a su significancia biológica. Estas bandas de frecuencia se denominan delta (δ), theta (θ), alfa (α), MU (μ), beta (β) y gamma (γ) de menor a mayor, respectivamente [65].

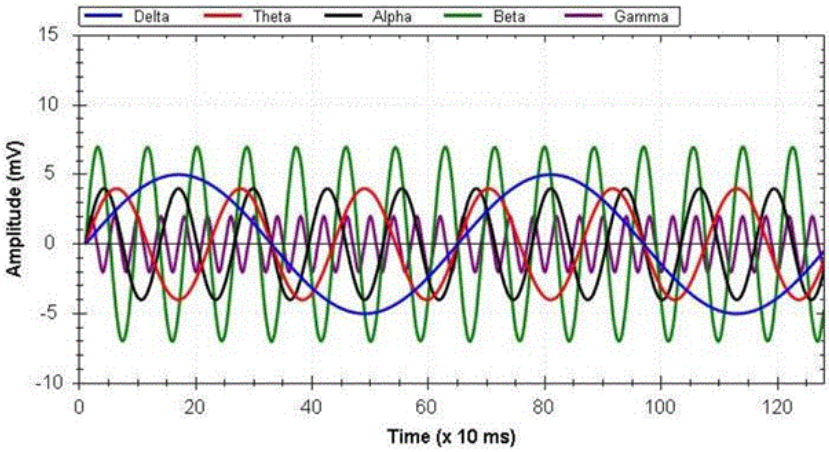

Figura 2.5: *Las principales bandas de frecuencia y su relación con las demás. Fuente: [65].*

Las ondas delta (δ) están presentes en niños menores de 1 año y en adultos en estado de sueño profundo. Una gran cantidad de actividad delta en adultos despiertos es anormal y se relaciona con enfermedades neurológicas. Debido a su baja frecuencia, es fácil de confundir las ondas delta con señales de artefactos que son causadas por los grandes músculos del cuello o la mandíbula. Este tipo de ondas no son útiles en los sistemas BCI [50].

Las ondas theta (θ) se localizan en la zona parietal-temporal del cerebro. Se producen por un estrés emocional, especialmente frustración o desilusión [121]. También se han asociado con el acceso al material inconsciente, inspiración creativa y meditación profunda. Altos niveles de ondas theta son consideradas anormales en adultos, y esta, por ejemplo, muy relacionada con AD/AH (*Déficit de Atención e Hiperactividad*) [52].

Las ondas alfa (α) están presentes durante estados de vigilia, sobre regiones posteriores de la cabeza, generalmente de gran amplitud sobre las zonas occipitales. La amplitud de las ondas alfa aumenta al cerrar los ojos o en estados de relajación y poca actividad mental y se atenúan o disminuyen cuando se abren los ojos, durante la atención, especialmente visual, y el esfuerzo mental [87].

Las ondas MU (μ) son asociadas con actividades motoras y son también encontradas en el rango de frecuencia de ondas alfa, pero su máxima amplitud es grabada sobre la corteza motora. De esta forma, básicamente se desencadenan cuando hay un movimiento actual o hay un intento de movimiento [21].

Las ondas beta (β) están a menudo divididas en $\beta 1$ (ondas beta bajas: 13,5 a 17 Hz), $\beta 2$ (ondas beta: 17,5 a 21 Hz) y $\beta 3$ (ondas beta altas: 21,5 a 30 Hz) para tener un rango más específico. Son ondas pequeñas y rápidas, asociadas con la concentración y a actividades mentales intensas y están mejor definidas en las áreas centrales y frontales del cerebro [122].

Las ondas gamma (γ) son las ondas más rápidas del cerebro, se asocian con una mayor actividad mental (no son los pensamientos normales o resolución de problemas) y pueden incluir destellos de brillantez y explosiones repentinas de percepciones/intuiciones y momentos de extrema atención y concentración.

Las ondas gamma y beta juntas han sido asociadas con la atención, la percepción y la cognición [93]. Este tipo de ondas son menos utilizadas en BCI basados en EEG, porque los artefactos tales como la Electromiografía (*Electrom-*

yography, EMG) o Electrooculografía (*Electrooculography*, EOG) les afecta. Sin embargo, este rango está atrayendo creciente atención en investigación BCI porque, en comparación con beta tradicional y las señales de alfa, la actividad gamma puede aumentar la tasa de transferencia de información y ofrecer mayor especificidad espacial [84].

La Figura 2.6 muestra los rangos de frecuencias de las diferentes bandas (δ, θ, α, μ, β, γ) presentes en la señal EEG.

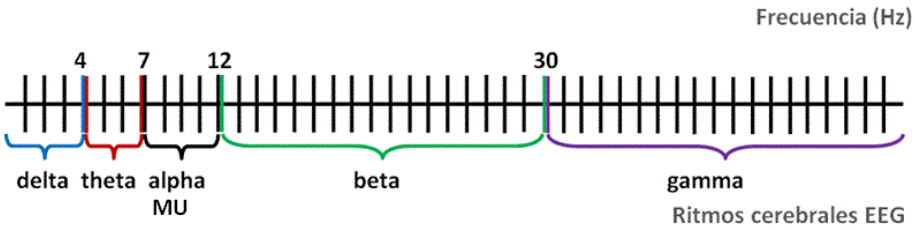

Figura 2.6: *Rangos de frecuencias de las diferentes bandas presentes en la señal EEG. Los rangos de la banda delta están por debajo de 4 Hz y los rangos de la banda theta oscilan entre 4 y 7 Hz. La mayoría de los sistemas BCI utilizan componentes de la banda alfa, cuyos rangos están entre 7 y 12 HZ como la banda MU y la banda beta entre 12 y 30 Hz. La banda gamma, que está empezando a aplicar en BCI, es mayor que 30 Hz.*

2.4.1.2. Potenciales Relacionados con Eventos

Los Potenciales Relacionados con Eventos (*Event Related Potential*, ERP) o Potenciales Evocados (*Evoked Potential*, EP) se definen como la respuesta neuroeléctrica cerebral específica, generada por la activación sincrónica de una población neuronal determinada, mediante un estímulo sensorial, motor o cognitivo interno o externo al sujeto, que activa selectivamente la región que se desea estudiar.

Los ERPs son respuestas de muy pequeña amplitud superpuestas y enmascaradas por la actividad de EEG, que es de mucha mayor amplitud, por lo que es necesario promediar múltiples registros con el estímulo desencadenante para que la respuesta sincronizada sobresalga de la actividad de base.

Los ERPs se pueden definir por la polaridad y latencia de sus componentes, con la sigla de la polaridad P si es positiva y N si es negativa, seguida de la latencia o retardo posterior a la estimulación en la que aparece dicho

componente, como por ejemplo, el potencial evocado P300 que describe una componente de polaridad positiva que aparece 300 ms después de una estimulación [16].

En relación a los sistemas BCI, los ERPs han marcado las líneas de investigación y en particular los potenciales evocados P300, los ritmos sensoriomotores μ y β, potenciales evocados visuales (VEP) y potenciales corticales lentos (SCP) [33]. Más adelante en la Sección 2.5 se hablará en detalle de los diferentes ERPs.

2.4.2. Magnetoencefalografía

Los diseños de BCI basados en Magnetoencefalografía (*Magnetoencephalography*, MEG), usa mediciones de los campos magnéticos producidos por la actividad eléctrica cerebral. Este tipo de medida directa de la actividad neuronal permite obtener mapas de la actividad cerebral con una alta resolución espacial y una alta resolución temporal del orden de los milisegundos. Esta excelente combinación espaciotemporal convierte a la MEG en una técnica ideal para el estudio de las funciones cognitivas.

Una ventaja de MEG sobre EEG es que los campos magnéticos son menos distorsionados por el cráneo y el cuero cabelludo que los campos eléctricos [15]. Sin embargo la mayor limitación para su amplia aceptación en BCI es el alto costo y el alto grado de tecnificación asociada con el equipo de MEG que usa los sensores "superconductores" [32].

Un enfoque representativo es la integración EEG-MEG [51]. El EEG hace su contribución desde los componentes radiales y tangenciales de la corriente primaria. De otra parte, las mediciones de MEG son insensibles al volumen de los efectos conductores, y son más sensibles a las fuentes tangenciales superficiales.

La Figura 2.7 representa las diferencias entre EEG y MEG, observándose que mientras que EEG detecta la diferencia de potencial eléctrico entre dos electrodos, MEG detecta los campos magnéticos dinámicos derivados de sus actividades eléctricas. De esta forma, los aspectos complementarios de las mediciones bioeléctricas y biomagnéticas sugieren que la combinación de EEG y MEG pueden proporcionar mejor confiabilidad y exactitud.

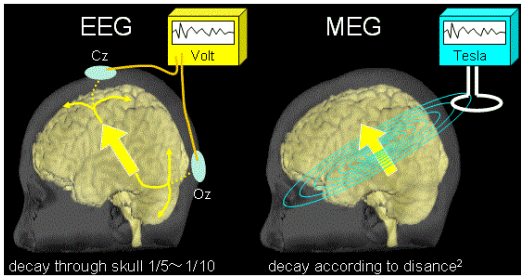

Figura 2.7: *Diferencias en las bases físicas entre EEG y MEG. Fuente: [5].*

2.4.3. Electrocorticografía

La Electrocorticografía (*Electrocorticogram*, ECoG) como plataforma de señal se ha convertido en los últimos años en un posible candidato para los sistemas BCI clínicamente viable.

ECoG es una técnica que mide la actividad eléctrica en la corteza cerebral por medio de electrodos profundos y electrodos subdurales que se colocan de forma intracraneal en el individuo. En comparación con EEG, ECoG proporciona una mayor resolución temporal y espacial y tiene una menor vulnerabilidad a los artefactos tales como EMG o movimientos oculares. Las señales registradas por medio de ECoG se encuentran en el rango de los mV y tienen buena SNR, lo que provoca que los usuarios necesiten una fase de entrenamiento mucho menor y a la hora de clasificar se obtienen tasas de acierto bastante mejores que usando EEG.

Sin embargo, los principales problemas que plantea esta técnica son el requerimiento de una craneotomía (operación quirúrgica del cráneo) para implantar una rejilla de electrodos, lo que implica riesgos significativos para la salud y que la vida útil de los electrodos implantados es limitada. Por esa razón, los primeros estudios sobre ECoG fueron realizados con animales. En la Figura 2.8, se muestra un ejemplo de ECoG mediante la utilización de una macro red de electrodos.

En sistemas BCI basados en ECoG, se ha demostrado, como un usuario logra controlar un cursor en una y en dos dimensiones [97]. Los resultados de estos estudios podrían hacer que sea más factible para las personas con discapacidades motoras severas utilizar BCI basados en ECoG para sus necesidades de comunicación y control.

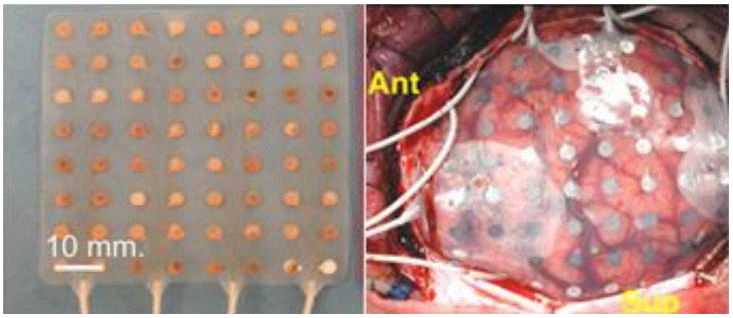

Figura 2.8: *Electroencefalograma Invasivo mediante la utilización de una macro red de electrodos. Fuente: [67]*

2.4.4. Adquisición en neuronas intracorticales

La adquisición en neuronas intracorticales es una técnica invasiva que requiere la implantación de microelectrodos en el interior de la corteza para medir la actividad neuronal eléctrica de las señales producidas por una sola neurona o campos eléctricos producidos por varias neuronas muy cercanas entre sí. Esta técnica ofrece una resolución espacial y temporal mucho más alta que EEG, por ello, las señales intracorticales resultan ser mucho más sencillas de utilizar por los pacientes que las señales EEG. No obstante, existe el inconveniente de que se requiere estabilizar la calidad de la señal a lo largo del tiempo, ya que los microelectrodos tienden a estropearse con el paso de los días y es posible que se deba hacer una recalibración de la sensibilidad de los electrodos para el usuario [84].

Se han empleado señales de control corticales para diseñar un BCI que permiten a los usuarios controlar el movimiento de un cursor y la flexión de un dedo en una mano virtual [59].

2.4.5. Imagen por Resonancia Magnética Funcional

La Imagen por Resonancia Magnética funcional (*functional Magnetic Resonance Imaging*, fMRI) detecta cambios en el volumen, flujo y nivel de oxigenación de la sangre a través de campos electromagnéticos permitiendo mostrar en imágenes las regiones cerebrales que ejecutan una tarea determinada. fMRI es un técnica de neuroimagen no invasiva ya que no se requiere inyecciones de sustancia alguna.

Los sistemas BCI basados en fMRI tienen por un lado, un enfoque manipulativo, ya que el sujeto es entrenado para cambiar voluntariamente la

actividad cerebral en una región particular del cerebro (como una variable independiente) con el fin de observar cambios en el comportamiento y además, tiene en cuenta la perspectiva físico-fisiológica al incorporar un paradigma experimental con una respuesta neuronal (como la variable dependiente). Usando *neurofeedback* en EEG, los estudios de SCP reportaron efectos en el comportamiento del procesamiento léxico, acción motora y desempeño musical [101].

Los sistemas BCI basados en fMRI, tienen la ventaja de señalar una región localizada del cerebro, con una alta resolución espacial y una razonable resolución temporal [113]. La Figura 2.9 presenta un ejemplo de una plataforma BCI basado en fMRI. Un sistema BCI basado en fMRI es un sistema de

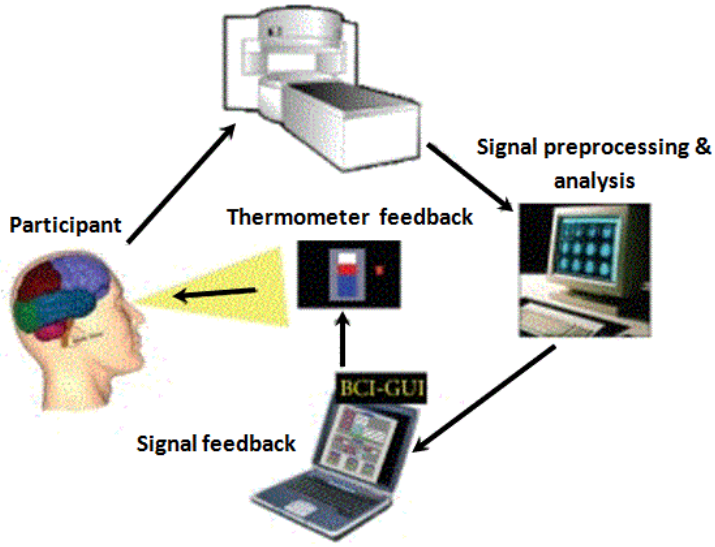

Figura 2.9: *Arquitectura general de un sistema BCI-fMRI y sus componentes. Fuente: [113].*

circuito cerrado que tiene los siguientes componentes principales: 1) adquisición de señal, 2) análisis de señal, 3) *feedback* de señal y 4) el participante.

Los primeros 3 componentes son usualmente ejecutados en computadores separados para optimizar el desempeño del sistema, y son conectados por una área de red local. La actividad cerebral espacialmente localizada es medida por el fMRI usando el efecto *"bold"* que es la respuesta neurovascular de la actividad eléctrica cerebral. Las series temporales de la señal de las regiones de interés son seleccionadas y después exportadas al software de visualización, el cual proporciona el *feedback* al sujeto usando proyecciones de video [113].

2.4.6. Espectroscopia del Infrarrojo Cercano

Los sistemas BCI basados en la Espectroscopia de Infrarrojo Cercano (*Near Infrared Spectroscopy*, NIRS) son relativamente nuevos y usan la región cercana del infrarrojo del espectro electromagnético (desde 800nm a 2500nm). La penetración es profunda dentro del cráneo dado que usa NIR; la luz NIR es emitida y la reflexión desde las células sanguíneas (específicamente el nivel de oxigeno en la hemoglobina) es usada como una medición del flujo sanguíneo. Este flujo sanguíneo denota activación de áreas cerebrales (activación deducida desde las imágenes con alta intensidad de reflexión). Este procedimiento es conocido como un método óptico, dado que NIRS produce imágenes de intensidad [90].

EEG tiene resolución espacial baja y es vulnerable al ruido eléctrico. En contraste, NIRS es más resistente al ruido eléctrico e impone menos restricciones al movimiento del cuerpo que el EEG y fMRI dado que el sensor-detector puede ser acomodado en el cráneo, pero su principal impedimento es la respuesta vascular lenta (en el orden de unos pocos segundos). También aunque la mayor energía NIR es reflejada, una pequeña porción de la energía NIRS puede ser absorbida por las células cerebrales y potencialmente puede ser perjudicial a largo plazo [120].

En un estudio reciente, Nagaoka y su grupo [3] desarrollaron un sistema de rehabilitación NIRS-BCI. En su estudio, fueron aplicados estímulos eléctricos correspondientes a la información de nervios craneales al musculo bíceps braquial del usuario mediante el establecimiento de umbrales en las señales medidas desde NIRS con el fin de causar el movimiento de la articulación del codo.

2.4.7. Tomografía por Emisión de Positrones

Los sistemas BCI basados en la Tomografía por Emisión de Positrones (*Positron Emission Tomography*, PET), están basados en la medicina nuclear, capaz de generar imágenes en 3D de procesos en el interior del cuerpo humano.

La técnica consiste en la detección y seguimiento de los positrones emitidos por una partícula radioactiva (radionúclido o trazador radioactivo) añadida a la glucosa y que se produce durante su metabolización por las neuronas en el cerebro. Durante la emisión de positrones se generan los rayos gama, que luego serán detectados por el escáner. Una computadora analiza los ra-

yos gama y utiliza la información para crear un mapa de imagen del órgano o tejido en estudio. La cantidad de radionúclidos concentrados en el tejido afecta el brillo con el que aparece el tejido en la imagen e indica el nivel de funcionalidad del órgano o tejido [90].

De esta forma, son detectados los cambios en el flujo sanguíneo debidos a una actividad eléctrica específica. Por ejemplo, en un sujeto pensando en mover su mano derecha, la actividad cerebral será detectada en el hemisferio izquierdo y viceversa. Esta actividad eléctrica detectada puede ser traducida para aplicaciones BCI.

Entre las ventajas de PET está la alta resolución espacial, en el orden de milímetros, aunque con una resolución temporal muy pobre. Entre las desventajas se encuentra el costoso equipamiento necesario para su realización, la necesidad de personal especializado y por supuesto los efectos radioactivos. En la Figura 2.10) se puede apreciar el equipamiento médico para la realización de PET.

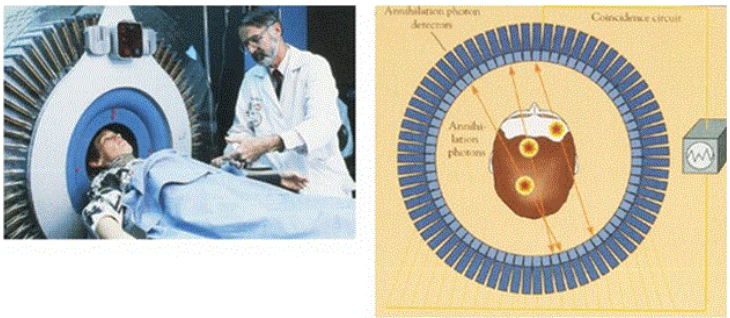

Figura 2.10: *Procedimiento para la realización de una PET. Fuente: [9].*

2.5. Tipos de BCI

Los sistemas BCI se pueden clasificar en dos grupos según la naturaleza de la señales de control empleadas: endógenos y exógenos [115].

2.5.1. Sistemas BCI endógenos

Los sistemas BCI endógenos no necesitan de ningún tipo de estimulación externa para la generación de la actividad cerebral necesaria para su funcionamiento y dependen de la capacidad del usuario para controlar determinadas características de su actividad electrofisiológica, como la amplitud

en una banda de frecuencia específica del EEG registrado sobre una zona concreta de la corteza cerebral [116]. Este tipo de sistemas son especialmente útiles para los usuarios con estados avanzados ALS o cuyos órganos sensoriales se han visto afectados.

Uno de los más importantes sistemas BCI de tipo endógeno, es el Dispositivo de Traducción de Pensamientos (*Thought Translation Device*, TTD) basado en la modulación voluntaria de los SCP, desarrollado en la Universidad de Tübingen [25].

Los sistemas endógenos requieren mucho entrenamiento para manejar un BCI y dependen significativamente del grado de modulación, muchas veces obtenidas a través de la Imaginación Motora (*Motor Imaginary*, MI) o de los ritmos cerebrales por del sujeto [44]. La atención sostenida y capacidad de predecir efectos futuros [82], frustración, cansancio y distracción [104], motivación [82], son algunos de los factores que pueden afectar a la capacidad para controlar las señales EEG, y han de ser tenidas en cuenta a la hora de establecer un protocolo de entrenamiento que garantice un aprendizaje más eficaz.

2.5.1.1. Potenciales Corticales Lentos

Los Potenciales Corticales Lentos (*Slow Cortical Potential*, SCP) se fundamentan en la teoría de la regulación voluntaria en el nivel promedio de EEG por parte del sujeto [99]. Los SCP son cambios de voltaje lentos que pueden apreciarse en intervalos superiores a los 5 segundos. Son generados en la corteza y tienen componentes espectrales entre 0.1 Hz a 1 Hz.

Se ha demostrado en muchos estudios, que los humanos somos capaces de adquirir el control de estos potenciales corticales (aprender a producir SCP en una dirección eléctricamente positiva o negativa) y otros ritmos de actividad cerebral (como la banda alfa) con la ayuda de técnicas de *biofeedback* adecuadas [91], [57]. A través de un TTD, es posible entrenar a los usuarios para que sean capaces de generar voluntariamente cambios en sus SCP.

Un TTD es un sistema de comunicación cerebro computador basado en EEG que ha sido desarrollado para restablecer la comunicación en pacientes con parálisis severa. Se basa en la auto-regulación de los SCP, la producción voluntaria de cambios de potenciales positivos y negativos. El TTD informa al usuario qué tipo de modulación ha realizado sobre sus SCP mediante un *feedback* visual o auditivo.

En un sistema BCI basado en SCP, se requiere que el sujeto regule su actividad cerebral de forma voluntaria, generalmente usando un paradigma de tipo binario en el que suele haber dos fases, la fase inicial de preparación seguida de la fase activa, durante la cual se le pide al sujeto que realice una actividad mental concreta (por ejemplo mover un cursor hacia arriba o hacia abajo de una pantalla de ordenador) basándose en estados emocionales o imaginación. En primer lugar, se mide el nivel de voltaje inicial durante la fase de preparación de unos 2 s y en los siguientes 3 s el usuario realiza la actividad mental concreta, produciendo una disminución o un aumento, respectivamente, de dicho nivel de voltaje.

En la Figura 2.11 se distinguen dos tipos de SCP: los SCP negativos, que están relacionados con la activación cortical, asociada a la realización de movimientos; y los SCP positivos, que están relacionados con la reducción de la activación cortical.

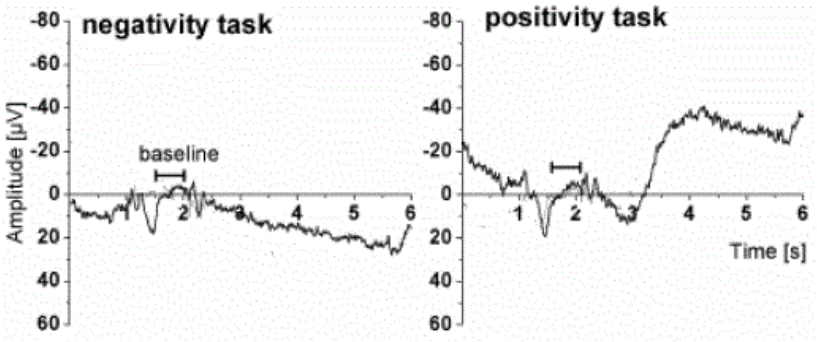

Figura 2.11: *La negatividad y la positividad del SCP en EEG. Fuente: [53].*

La autorregulación de los SCP se ha aplicado en la selección de letras dividiendo en cada intento el abecedario en dos partes, hasta escoger una sola letra. Se han alcanzado precisiones superiores al 75 % y tasas de comunicación de 0.15 a 3 letras por minuto [24]. El principal problema con este enfoque es que requiere un amplio entrenamiento, de hasta varios meses con 1 ó 2 días a la semana, de 7 a 12 ejecuciones por día y entre 70 y 100 ensayos por ejecución [53].

La producción de SCP positivos o negativos ha sido explorada por Birbaumer [24] para mostrar que los pacientes con ALS pueden utilizar un BCI para controlar un dispositivo al seleccionar letras de una pantalla y comunicarse así con su entorno.

Como se menciono antes, los sistemas BCI basados en SCP son por su naturaleza de tipo binario. Se han realizado intentos posteriores por conseguir un paradigma multidimensional en el que puedan usarse varios estados mentales, pero han resultado ser difíciles de controlar [62].

2.5.1.2. Ritmos Sensoriomotores μ y β

Los sistemas BCI basado en ritmos sensoriomotores, se basan en un paradigma de dos o más clases de imágenes motoras (movimiento de la mano derecha o izquierda, de los pies, de la lengua, etc.) u otras tareas mentales (rotación de un cubo, realización de cálculos aritméticos, etc.). Este tipo de tareas mentales producen cambios en la amplitud de los ritmos sensoriomotores μ (8-12 Hz) y β (13-28 Hz), registrados sobre la zona somatosensorial y motora de la corteza cerebral [54] [117]. Estos ritmos presentan variaciones tanto para la ejecución de un movimiento real como para la imaginación de un movimiento o la preparación al mismo.

Los ritmos cerebrales se han utilizado para proveer al usuario el control sobre diversos dispositivos que le permiten manejar su medio u operar programas en un ordenador. La Figura 2.12 muestra el fenómeno básico de los ritmos sensoriomotores. El EEG es grabado sobre la corteza sensoriomotora. Los usuarios controlan la amplitud del ritmo μ de 8-12 Hz (o de 13-28 Hz en el ritmo β) para mover un cursor a un objetivo en la parte superior de la pantalla o a otro objetivo en la parte inferior (u objetivos adicionales en localizaciones intermedias).

El espectro de frecuencia (arriba) para los objetivos de superior e inferior muestran que el control está claramente enfocado en la banda de frecuencia de ritmo μ. Muestras de trazos (abajo) de EEG también indican que el ritmo μ es prominente cuando el objetivo está en la parte superior y mínimo cuando está en la parte inferior de la pantalla [116].

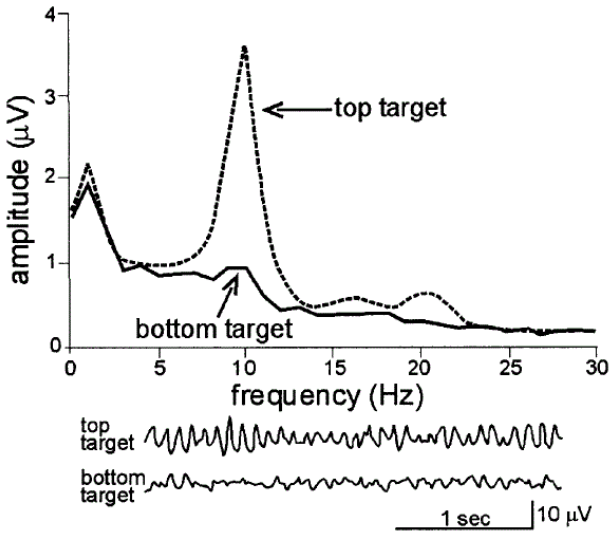

Figura 2.12: *Ritmos Sensoriomotores en BCI. Fuente: [116].*

Stecklow, Infantosi y Cagy en 2008 [102] hicieron un estudio con deportistas y no deportistas comparando la Imaginación Motora (*Motor Imaginary*, MI) de una tarea previamente entrenada. Observaron resultados diferenciados en esas señales, sin embargo, confirman la participación también de la corteza parietal posterior del hemisferio contralateral en la planificación y simulación de tareas mentales. También indicaron que esa lateralización en las diferentes señales (activación en el electrodo C3) ocurre según el conocimiento previo de MI que tenga el individuo.

2.5.2. Sistemas BCI exógenos

Los sistemas BCI exógenos dependen de la actividad electrofisiológica evocada por estímulos externos para producir una serie de respuestas fisiológicas que serán moduladas de forma voluntaria por el sujeto mediante tareas cognitivas. La modulación voluntaria de estas respuestas fisiológicas causa distintos patrones de actividad cerebral que serán usados en los sistemas BCI [116].

Los sistemas exógenos tienen la ventaja de que no requieren prácticamente entrenamiento por parte del sujeto. Normalmente, se presenta al sujeto de forma simultánea un conjunto de estímulos distintos y cada uno de éstos evoca una respuesta fisiológica distinta. Cada estímulo representa un comando o código. Mediante alguna actividad cognitiva, el sujeto es capaz de modular la respuesta fisiológica del estímulo objetivo y así se reconoce el código que

el sujeto quiere comunicar.

Los sistemas BCI de tipo exógenos son los basados en los Potenciales Evocados Visuales de Estado Estable (*Steady State Visual Evoked Potential*, SSVEP), potenciales P300 y el uso de un paradigma de tipo *oddball*, que se basa en el reconocimiento, de entre una serie de estímulos idénticos o no relevantes aquél distinto o de interés.

2.5.2.1. Potenciales Evocados Visuales

Los Potenciales Evocados Visuales (*Visually Evoked Potential*, VEP) son potenciales recogidos sobre la corteza visual, derivados de una respuesta cerebral a un estímulo visual [112]. Los VEP de Estado Estable (SSVEP) son respuestas periódicas causadas por la presentación repetida de una serie de estímulos visuales a una frecuencia superior de 6 Hz. La potencia de los SSVEP se extiende sobre un ancho de banda muy estrecho, ya que la frecuencia de la respuesta coincide con la de la excitación [58].

En [31] un BCI basado en SSVEP es usado para marcar un número de teléfono (ver Figura 2.13).

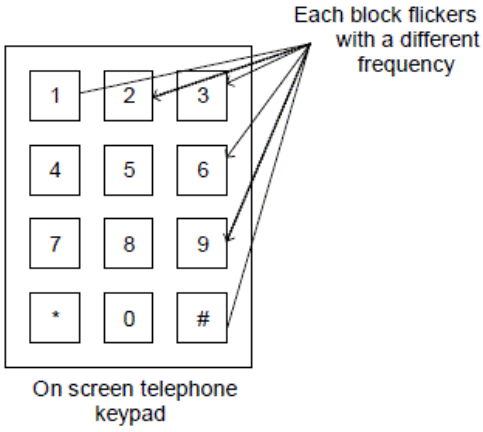

Figura 2.13: *Diseño de un sistema BCI basado en SSVEP usado para marcar un número de teléfono. Fuente: [90].*

Se muestra cada número como un estímulo visual que parpadea a una cierta frecuencia de forma que el usuario debe mantener fija la mirada en el

número que desea marcar. El sistema BCI calcula la potencia de todas las frecuencias de parpadeo y clasifica la que con más probabilidad corresponde al número atendido. De esta forma se produce un SSVEP en la corteza visual, que consiste en una componente periódica de la misma frecuencia a la que parpadea el número y una serie de armónicos secundarios.

La Figura 2.14 muestra dos SSVEP, a 8,5 Hz y 13 Hz, cada uno de ellos provocado por separado. La mejor respuesta se obtiene en el rango de 5-20 Hz [46]. Con este sistema se han obtenido altas Velocidades de Transferencia de Información (*Information Transfer Rate*, ITR), con un máximo cercano a 1 *bps*.

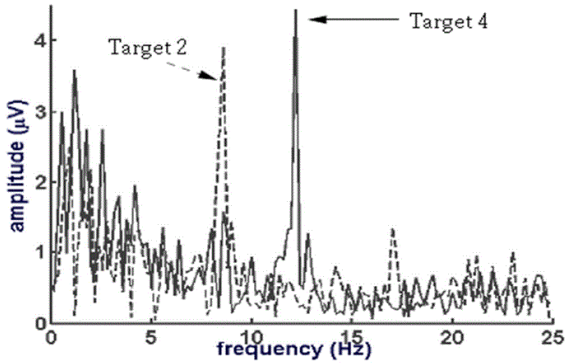

Figura 2.14: *Frecuencias SSVEP para dos objetivos. Fuente: [28].*

Los sistemas BCI basados en VEP tienen especial interés debido a la robustez de los fenómenos SSVEP, a la buena SNR, debido fundamentalmente a que el ancho de banda necesario para la extracción de la característica se extiende tan solo al de la frecuencia de repetición del estímulo permitiendo esto, obtener una ITR aceptable, dentro del campo de los BCI. Así mismo, no requieren entrenamiento, ya que los SSVEP se generan en la corteza mediante mecanismos fisiológicos y la modulación de estos se produce mediante la atención selectiva, que es una característica innata en el ser humano. Se han conseguido precisiones del 84 % con estímulos a 6 frecuencias diferentes [46], así como tasas de comunicación de hasta 10-12 palabras por minuto [112].

2.5.2.2. Potenciales Evocados P300

El potencial P300 es un pico de amplitud que aparece en el EEG aproximadamente unos 300 ms después de haberse producido un estímulo auditivo o visual poco frecuente. Habitualmente, se presenta al usuario una tanda de estímulos de los que solo unos pocos tienen relación con la intención del

usuario. De esta forma, los estímulos de interés, al ser infrecuentes y estar mezclados con otros estímulos mucho más comunes, provocan la aparición de un potencial P300 en la actividad cerebral del usuario [54]. Dicho potencial se observa principalmente en las zonas central y parietal del la corteza cerebral [60].

La amplitud del P300 varía directamente con la relevancia del estímulo provocado y con la habilidad del sujeto para discriminar unos eventos de otros.

La primera aplicación BCI basada en potenciales evocados P300 fue implementada por Farwell y Donchin en 1988 [42], también llamado paradigma P300. Consiste en una matriz de 6 filas y 6 columnas que contiene letras y números del alfabeto, con el propósito de mandar mensajes escritos. Se obtuvo un ITR de 12 *bpm* con sujetos sanos (ver la Figura 2.15).

Figura 2.15: *Matriz de letras formada por 6 filas y 6 columnas de letras que se van iluminando aleatoriamente. Fuente: [36].*

Las filas y las columnas de esa matriz se iluminan aleatoriamente. Los sujetos deben prestar atención a una de las letras o números de la matriz y contar cuántas veces se ilumina. Tras doce iluminaciones, cada fila y columna se ha iluminado una vez y, por tanto, el elemento objetivo del usuario se ha iluminado dos veces. Se promedia, entonces, la respuesta para cada elemento de la matriz y el potencial que presente un pico de mayor amplitud será el que el usuario pretende seleccionar. La Figura 2.16 muestra la forma de P300 promediado.

Las características fundamentales de este tipo de BCI es que necesitan de muy poco entrenamiento, no causa mucho cansancio y es adecuado para pacientes que no pueden controlar ningún movimiento voluntario del sistema motor, debido a que se basa en habilidades cognitivas y sensoriales de los usuarios [92].

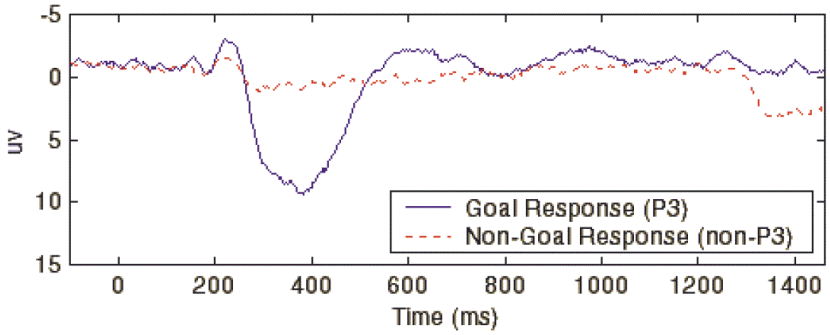

Figura 2.16: *P300 promediado. La línea continúa muestra un P300 evocado cuando el sujeto cuenta el número de veces que un estímulo visual le es presentado. La línea punteada muestra la onda típica cuando el recuento no se produce ante la misma estimulación. Fuente: [19].*

Una versión desarrollada con posterioridad del paradigma P300 mostró ser algo más eficiente con sujetos con deficiencias motoras [36]. En [95] Rebsamen presenta al usuario una pantalla con diferentes opciones que se corresponden con localizaciones de un espacio público.

En [19] se muestran dos estudios de BCI con P300 en entornos de realidad virtual. El estudio de Iturrate en el que además de controlar los movimientos de la silla de ruedas en tiempo real mediante los potenciales evocados P300, el usuario esta remotamente conectado vía internet con una estación robot que puede estar localizada en cualquier parte del mundo [40].

Y por último destacar el BCI desarrollado en [86] ya que está basado en estímulos auditivos en vez de visuales. La idea es que pueda ser usado con sujetos en estado muy avanzado de ALS, cuando el sujeto tiene limitaciones para dirigir la mirada o incluso enfocar correctamente un estímulo visual.

2.6. Etapas consecutivas de un BCI

El funcionamiento de un sistema BCI se puede dividir en cinco etapas como se muestra en la Figura 2.17: (1) adquisición de la señal, (2) preprocesamiento o mejora de la calidad de la señal, (3) extracción de características, (4) clasificación y (5) control del dispositivo externo. En algunos documentos científicos, se suele encontrar una etapa denominada procesamiento que integra la etapa de extracción y la etapa de clasificación de características.

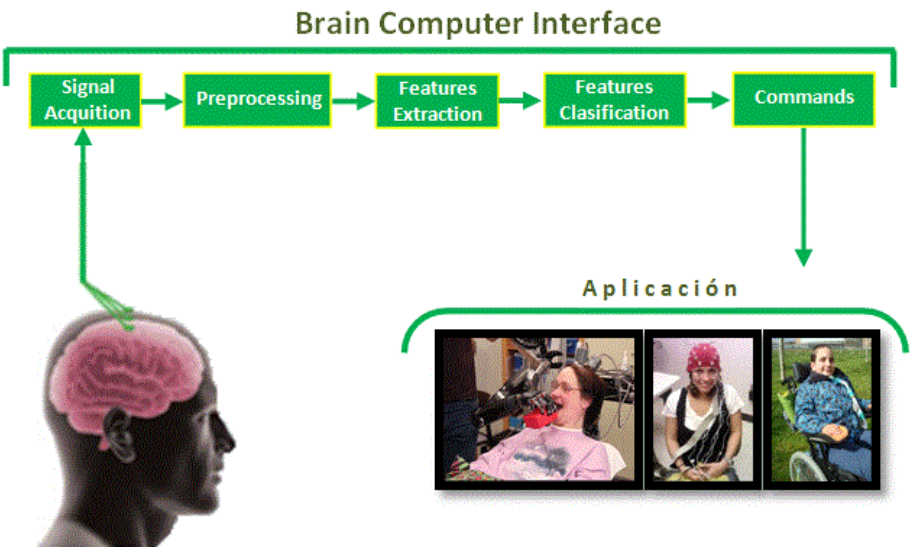

Figura 2.17: *Etapas consecutivas de un sistema BCI.*

2.6.1. Etapa de adquisición de la señal

La etapa de adquisición de la señal captura el fenómeno neurológico que refleja las intenciones del usuario mediante electrodos, que son además amplificadas, filtradas y digitalizadas. También puede llevar a cabo la reducción de ruido y el procesamiento de artefactos tales como EMG y EOG, Electrocardiograma (*Electrocardiography*, ECG) y Respuesta Galvánica de la Piel (*Galvanic Skin Response*, GSR) como medidas objetivas sobre el estado del sujeto.

En esta primera etapa se prepara la señal registrada para su posterior procesado. Aunque para el procesado *online* y, en consecuencia, para el funcionamiento del sistema BCI no resulta necesario almacenar la señal registrada, casi todos los sistemas BCI incorporan esta etapa con objeto de permitir posteriores análisis y procesados de la misma.

2.6.2. Etapa de pre-procesado

La mayoría de las señales biomédicas aparecen como señales débiles en un ambiente que está lleno con muchas otras señales de diversos orígenes. Cualquier señal que no sea el de interés podría ser denominada como una interferencia, artefacto o simplemente ruido. Las fuentes de ruido pueden ser fisiológicas, la instrumentación utilizada o el medio ambiente del experimento [94].

En esta etapa se utiliza diferentes tipos de diseños de filtros que abarcan una amplia gama de enfoques así como algoritmos para eliminar de la señal en bruto los artefactos no deseados. Los filtros permiten eliminar frecuencias no deseadas que contaminan y dificultan la interpretación de la señal. Esto debe hacerse procurando no sustraer aquellas frecuencias que forman parte esencial de la actividad bioeléctrica cerebral. La etapa de pre-procesado prepara las señales adquiridas en la etapa anterior en una forma adecuada para su posterior procesamiento.

2.6.3. Etapa de extracción de características

La extracción de características permite obtener determinadas características específicas de la señal adquirida que son útiles para discriminar entre diferentes tareas mentales. Se crea una serie de vectores de propiedades que caracterizan a lo largo del tiempo las cualidades de la señal adquirida.

El diseño de un conjunto adecuado de características es una cuestión difícil ya que la información de interés de las señales del cerebro se oculta en un entorno muy ruidoso, son inherentemente no estacionarias y comprenden un gran número de fuentes simultáneas. Una señal que puede ser de interés podría ser superpuesta en el tiempo y en el espacio por múltiples señales desde tareas cerebrales diferentes. Se requiere además, que la dimensión del vector de propiedades no sea demasiado elevada con el fin de no incrementar demasiado la complejidad del sistema y que esta reducción del número de propiedades no signifique una pérdida de información relevante de la señal [55], [84].

2.6.3.1. Métodos de extracción de características

Los métodos para la extracción de características consisten en la realización de distintas combinaciones y transformaciones sobre las señales cerebrales recogidas en la etapa de adquisición. A partir de una o varias carac-

terísticas obtenidas de esta manera se realizará posteriormente la etapa de clasificación.

La extracción de características se puede hacer ya sea en el dominio del tiempo, en el dominio de la frecuencia, y/o en el dominio de tiempo-frecuencia. Emplear uno u otro método va a depender de las señales de control con las que se utilice y de las características que se quieran resaltar. Los métodos para extraer características como Análisis de Componentes Principales (PCA), Análisis de Componentes Independientes (ICA) y Patrón Espacial Común (CSP) permiten encontrar una representación lineal de los datos para obtener la información más relevante. Para el análisis de los datos en tiempo-frecuencia, se usan métodos como la Transformada de Fourier por intervalos (STFT), Transformada Wavelet (WT), Modelos autorregresivos (AR) y Filtro Adaptado (MF) con el mismo objetivo.

Todos estos métodos se explican a continuación a excepción de PCA, WT y Modelos AR, que se explican más adelante en el Capítulo 4, sección 4.2, ya que son los métodos empleados para extraer las características de los datos de estudio de SCP en este TFM.

2.6.3.1.1. Análisis de Componentes Independientes

Análisis de Componentes Independientes (*Independent Component Analysis*, ICA), es un método estadístico que permite encontrar una representación lineal de los datos de forma que las componentes sean estadísticamente lo más independientes posible. Una representación de este tipo permite obtener la estructura fundamental de los datos en muchas aplicaciones, incluidas la extracción de características y la Separación Ciega de Fuentes (*Blind Source Separation*, BSS) [100].

ICA establece 3 hipótesis: (1) las fuentes son estadísticamente independientes entre sí, (2) las fuentes tienen distribución no gaussiana y (3) el modelo de mezcla se conoce (lineal o no lineal). Una vez obtenidos los datos, estos se conocen como las componentes independientes.

El clásico ejemplo para explicar el funcionamiento de ICA, es el llamado problema del cóctel de fiesta [103], representado en la Figura 2.18, donde estas en un cuarto con varias personas hablando al mismo tiempo, de tal manera que solo se puede escuchar la mezcla de voces y es muy difícil identificar lo que cada persona está diciendo. Pensando que las fuentes provienen

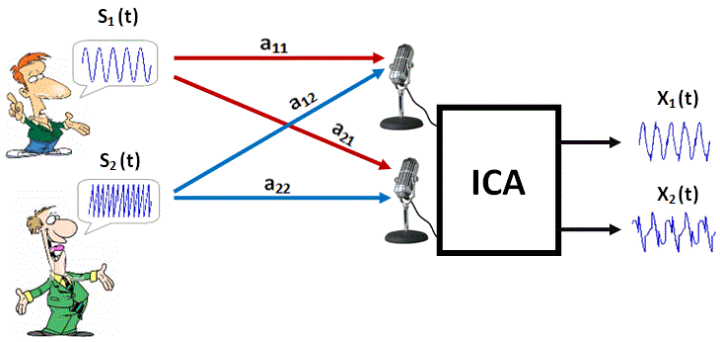

Figura 2.18: *Representación de ICA: problema del cóctel de fiesta. Fuente: [110].*

de dos voces del cóctel de fiesta y se tuvieran dos micrófonos en diferentes lugares, cada uno registraría una señal en el tiempo, las cuales, se pueden escribir como $x_1(t)$ y $x_2(t)$. Cada una de esas señales registradas (señales observadas) es una suma ponderada de las señales emitidas por los dos locutores (señales originales), las cuales denotamos como $S_1(t)$ y $S_2(t)$. Lo cual, se puede expresar como una transformación lineal:

$$\begin{aligned} x_1(t) &= a_{11}S_1(t) + a_{12}S_2(t) \\ x_2(t) &= a_{21}S_1(t) + a_{22}S_2(t) \end{aligned} \tag{2.1}$$

o de forma análoga:

$$X = A * S \tag{2.2}$$

donde los coeficientes a_{ij} son constantes que dan los pesos de la mezcla y que se agrupan en la matriz A, (en otras palabras, los parámetros a_{ij} dependen de las distancias de los micrófonos a los locutores), $x_i(t)$ representa las señales observadas y $S_i(t)$, son los datos de las señales originales.

El problema ahora es estimar las señales originales $S_1(t)$ y $S_2(t)$ y la matriz de mezcla A contando únicamente con las señales mezcladas $x_1(t)$ y $x_2(t)$.

Los coeficientes de mezclado a_{ij} toman una serie de valores que permiten que la matriz A sea invertible. De este modo, existe una matriz W con coeficientes w_{ij} tal que permite separar las señales $S_i(t)$ como:

$$\begin{aligned} S_1(t) &= w_{11}x_1(t) + w_{12}x_2(t) \\ S_2(t) &= w_{21}x_1(t) + w_{22}x_2(t) \end{aligned} \tag{2.3}$$

o en forma matricial:

$$S = W * X \tag{2.4}$$

Entonces se encuentra la matriz W como la inversa de A, la matriz que formaban los coeficientes a_{ij}, si se conocen dichos coeficientes de mezclado.

La Figura 2.19, ilustra el esquema que representa los procesos de generación de observaciones y estimación de fuentes en el problema de BSS.

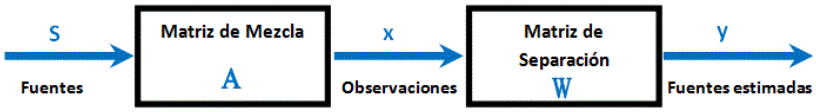

Figura 2.19: *Esquema general que representa los procesos de generación de observaciones y estimación de fuentes en el problema de BSS.*

ICA obtiene las señales originales $S_i(t)$ y la matriz de mezcla A de las señales observadas $x_i(t)$, por medio de algoritmos tales como Infomax. Dichos algoritmos son rápidos y robustos a pesar de su poca flexibilidad [20].

En los sistemas BCI, ICA es empleado usualmente para descomponer los datos de las señales cerebrales de EEG eliminando los artefactos oculares que están inmersos en dichas señales.

2.6.3.1.2. Patrones Espaciales Comunes

Los Patrones Espaciales Comunes (*Common Spatial Pattern*, CSP) es un método de extracción de características que ha sido empleado con gran éxito para el cálculo de filtros espaciales empleados para la detección de ERD y ERS en los sistemas BCI [84]. CSP es un método de filtrado espacial que procesa los datos de entrada para que tengan una varianza óptima para su posterior clasificación. Se ha empleado CSP en datos que proceden de múltiples fuentes o canales y que puede pertenecen tanto a dos clases diferentes como a múltiples clases.

2.6.3.1.3. Transformada de Fourier

La Transformada de Fourier (*Fourier Transform*, FT) permite descomponer una señal en sus componentes sinusoidales de diferentes frecuencias, en

otras palabras, puede pensarse que es una técnica matemática para transformar el punto de vista de una señal desde la base de tiempo a la base de la frecuencia, tal como se representa esquemáticamente en la Figura 2.20.

Figura 2.20: *Esquema de la Transformada de Fourier. Fuente: [75].*

El análisis mediante la FT resulta extremadamente útil en muchos casos, razón por la cual sería natural preguntarse por qué podría ser necesario el uso de otra herramienta para el análisis de señales; pues bien, al pasar una señal al dominio de la frecuencia se pierde la información referente al tiempo; más precisamente, cuando se observa una señal producto de la FT, resulta imposible determinar cuándo ocurre un determinado evento o cuándo está presente una determinada frecuencia [75].

Si la señal que se está analizando es estacionaria, ésta desventaja no resulta muy relevante. Sin embargo, un importante número de señales de interés presentan características no estacionarias o transitorias como el EEG y la FT no está preparada para detectarlas y/o analizarlas [75].

Matemáticamente, la FT se define como:

$$F(t) = \int_{-\propto}^{\propto} y(t).e^{-j2\pi ft}dt \tag{2.5}$$

donde *y(t)* es la señal en el dominio del tiempo, *F(t)* es la señal en el dominio de la frecuencia, i es la unidad imaginaria equivalente a $\sqrt{-1}$ -1 y $e^{-j2\pi ft}$ es la función exponencial compleja, que consiste en la combinación de una función seno y una función coseno (o, lo que es lo mismo, en la combinación de dos funciones sinusoidales que presentan un desfase entre ellas de $\pi/2$).

En un esfuerzo por corregir la deficiencia presentada con la FT, en 1946 Denis Gabor adaptó la Transformada de Fourier para poder analizar una pequeña sección de la señal en un determinado tiempo (mediante una especie de ventana). Esta adaptación es la que se conoce como Transformada de Fourier por intervalos (*Short-Time Fourier Transform*, STFT), la cual lleva

una señal del plano del tiempo al plano bidimensional de tiempo y frecuencia, tal como se presenta esquemáticamente en la Figura 2.21.

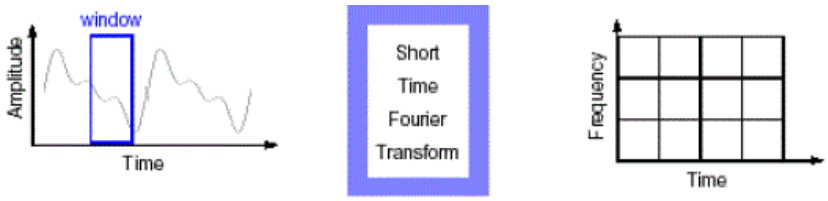

Figura 2.21: *Esquema de la Transformada de Fourier por intervalos (STFT). Fuente: [75].*

La STFT representa una especie de compromiso entre el dominio del tiempo y el de la frecuencia de una señal, ya que provee algo de información acerca de cuándo y a qué frecuencia de una señal ocurre un determinado evento. Sin embargo, solamente se puede obtener dicha información con una precisión limitada, la cual está acotada por el tamaño de la ventana. Mientras que el compromiso entre la información del tiempo y la frecuencia puede resultar útil, el inconveniente surge dado que una vez que se escoge un determinado tamaño para la ventana de tiempo, dicha ventana es la misma para todas las frecuencias. Muchas señales requieren un acercamiento más flexible, de modo tal que sea posible variar el tamaño de la ventana para determinar con mayor precisión el tiempo o la frecuencia [75].

2.6.3.1.4. Filtros Adaptados

Los filtros Adaptados (*Matched Filtering*, MF) o de correlación es un método de extracción de características en el que un tipo de filtro se diseña con el propósito de detectar un patrón específico en las señales cerebrales en base a su similitud con un conjunto de señales o plantillas conocidas predeterminadamente. La intención del usuario es revelada por medio de la correlación entre las señales EEG desconocidas y el conjunto de plantillas. Cada plantilla que pertenece a ese conjunto de plantillas representa una intención concreta del usuario y está representada por un filtro que se emplea para calcular la correlación entre la plantilla y la señal. Una mayor correlación implica una mejor adecuación entre la plantilla y la intención del usuario con lo que existe mayor probabilidad de que efectivamente el usuario pretendía generar ese patrón en su señal EEG.

Los MF han sido utilizados para la extracción de características de los SSVEP, así como para identificar las intenciones de un usuario a través de

ritmos μ [55], [84].

2.6.4. Etapa de clasificación de características

Si en la etapa anterior se trataba de obtener ciertos parámetros que caracterizaran las señales EEG procedentes del sujeto, en esta etapa se pretenden clasificar dichos parámetros entre diferentes patrones o clases. El éxito de la clasificación viene determinado por la apropiada elección de los parámetros que caracterizan la señal, por la efectividad de los propios algoritmos de clasificación y por supuesto, por la capacidad del sujeto para producir o controlar dichas características. Si un sujeto no tiene control, los algoritmos de clasificación no podrán asociar parámetros a estados mentales deseados y el sistema BCI no funcionará.

En la actualidad los sistemas BCI se pueden clasificar entre más de dos clases, sin embargo, cuanto mayor sea el número de éstas, peores serán las prestaciones del sistema. Distinguir entre un número elevado de estados mentales implica el uso de algoritmos de clasificación especialmente robustos y un absoluto control por parte del sujeto para producir diferentes patrones EEG, lo que en la actualidad está lejos de conseguirse.

En los sistemas BCI, la elección del algoritmo de clasificación viene condicionado principalmente por el número de clases y por la posibilidad de adaptación al sujeto.

2.6.4.1. Métodos de clasificación de características

Hay varias metodologías populares de clasificación que han sido exploradas en BCI: clasificadores lineales y clasificadores no lineales [90].

Los clasificadores lineales suelen ser más robustos que los no lineales [90]. Estos algoritmos asumen que los datos pueden separarse de manera lineal. Divide el espacio en hiperplanos de separación generando diferentes regiones. Cada una de estas regiones se corresponde con una selección distinta. Los algoritmos no lineales tienen en cuenta más parámetros por lo que la carga computacional es mayor. Cuando se trabaja con pocos datos los algoritmos lineales dan buenos resultados. Si se trabaja con grandes cantidades de datos son más adecuados los métodos no lineales [78].

En este TFM se estudian dos métodos para el diseño de la etapa de clasificación de un BCI, el Análisis Discriminante Lineal y una Máquina de

Vectores de Soporte que se explican más adelante en detalle en el Capítulo 4, sección 4.3.

2.6.4.1.1. Red Neuronal Artificial

Una Red Neuronal Artificial (*Artificial Neural Network*, ANN), es un clasificador no lineal que se basa en una analogía del sistema nervioso. La idea general consiste en emular la capacidad de aprendizaje del sistema nervioso, de manera que la ANN aprenda a identificar un patrón de asociación entre los valores de un conjunto de variables predictoras (entradas) y los estados que se consideran dependientes de dichos valores (salidas) [45].

Una ANN contiene un conjunto de elementos simples de procesado (nodos) interconectados. La capacidad de procesado de la red reside en los valores asociados a las conexiones entre nodos, denominadas pesos, que se obtienen en un proceso de adaptación o aprendizaje a partir de un conjunto de patrones de entrenamiento.

Matemáticamente, una ANN define una aplicación de un espacio de entrada a un espacio de salida que puede ser descrito como una función vectorial $Y = f(x)$, donde X e Y pueden tener cualquier dimensión. Las ANN se componen de un conjunto de nodos o neuronas artificiales que procesan la información por medio de la función no lineal.

Existen distintos tipos de redes neuronales, algunos muy sencillos, con una sola neurona (perceptrón), hasta redes que se componen por decenas de ellas. El perceptrón tiene la capacidad de clasificar muestras que sean linealmente separables, sin embargo, si se necesita una clasificación más compleja, donde la separación de las clases solo se puede hacer mediante varias funciones discriminantes entonces se necesita una red neuronal multicapa, o también conocida como el Perceptrón Multicapa (*Multilayer Perceptron*, MLP).

MLP es uno de los clasificadores no lineales más comúnmente empleados en diseños de BCI. MLP está formada por múltiples capas (esto le permite resolver problemas que no son linealmente separables) que contienen varias neuronas artificiales (*perceptrones*) que realizan distintas transformaciones sucesivas.

Un ejemplo de este tipo de red neuronal se muestra en la Figura 2.22. Este tipo de red neuronal, normalmente tiene tres capas activas, una capa de sa-

lida y dos capas ocultas. La capa de entrada no contribuye en el algoritmo y solamente transmite su entrada a la capa siguiente. En MLP las señales van siempre desde la capa de entrada a la capa de salida.

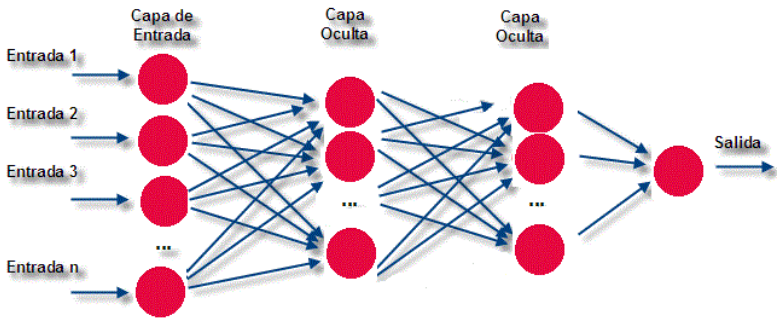

Figura 2.22: *Estructura de un MLP con una capa de entada y tres capas activas. Fuente: [10].*

Matemáticamente, la salida generada por un MLP se puede describir como una suma ponderada de varias entradas:

$$y = f\left(\sum_{i=1}^{N} w_i x_i - w_0\right) \tag{2.6}$$

donde y es el término de salida, x_i es la i-ésima entrada de las N entradas y w_i su peso asociado. La función f es siempre una función monótona y, por lo general, es la función de Heaviside, $H(z)$, o la función sigmoidea logística $\varphi(z)$.

$$H(z) = \begin{cases} 1 & z \geq 0 \\ 0 & z < 1 \end{cases}$$

$$\varphi(z) = \frac{1}{1 + e^{-\beta z}} \tag{2.7}$$

Ambas funciones comparten la propiedad de que sus resultados tienden a cero para entradas negativas y a uno para entradas positivas. Esto significa que las funciones conmutarán de cero a uno cuando la suma ponderada de las entradas sea mayor que el término umbral w_0.

El valor de cada peso w_i es modificado durante el proceso de aprendizaje. En el aprendizaje, se introducen varias entradas a la ANN y se analiza la

salida. Si la salida es incorrecta se ajustan internamente los pesos para reducir el riesgo de cometer el mismo error en la siguiente. Con esta modificación se espera que la ANN responda correctamente la próxima vez cuando tenga las mismas o similares entradas (aprendizaje supervisado).

2.6.4.1.2. K-vecinos más Cercanos

K-Vecinos más Cercanos (*Nearest Neighbour*, K-NN) es un modelo clasificador de mínima distancia y se ubica en los clasificadores basados en métricas. El paradigma de K-NN se fundamenta por tanto en una idea muy simple e intuitiva, lo que unido a su fácil implementación hace que sea muy extendido.

Para hacer la clasificación, K-NN utiliza un tipo de distancia, con la que determina todas las distancias entre el punto a clasificar y todos los puntos del conjunto de entrenamiento. Con las distancias calculadas determina los K vecinos más cercanos y, según el tipo de la clase para determinar, asigna el punto a una de ellas.

La distancia es el criterio de comparación principal en K-NN. La distancia clásica Euclidiana es comúnmente el tipo de métrica más utilizada. Existen métricas alternativas tales como, la distancia de Manhattan, distancia de Chebychev, distancia de Mahalanobis, entre otras.

La distancia euclidiana, es la distancia entre dos puntos $p1$ y $p2$ en un plano cartesiano, esto es:

$$d(p1, p2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{2.8}$$

En K-NN, la letra K, es usualmente un número entero que indica el número de patrones que se consideran para tomar la decisión; así, se habla del clasificador 1-NN o del 3-NN.

La Figura 2.23 muestra 26 casos ya clasificados en dos posibles valores: la clase 1 (círculos rojos rellenos) y la clase 2 (círculos rojos sin relleno). Las variables predictoras son X_1 y X_2. La muestra que se desea clasificar es el triángulo verde. Para $K = 3$, este es clasificado con la clase 2, ya que hay solo un círculo rojo relleno y dos círculos rojos sin relleno, dentro del círculo interno que los contiene y por tanto el clasificador 3-NN predice la clase 2 para el nuevo caso. Nótese que el caso más cercano al triángulo verde perte-

nece a la clase 1, si hubiésemos utilizado un clasificador 1-NN, el triángulo verde hubiese sido asignado a la clase 1.

Si $K = 5$, este es clasificado con la clase 1, ya que hay 2 dos círculos rojos sin relleno y 3 círculos rojos rellenos, dentro del círculo externo.

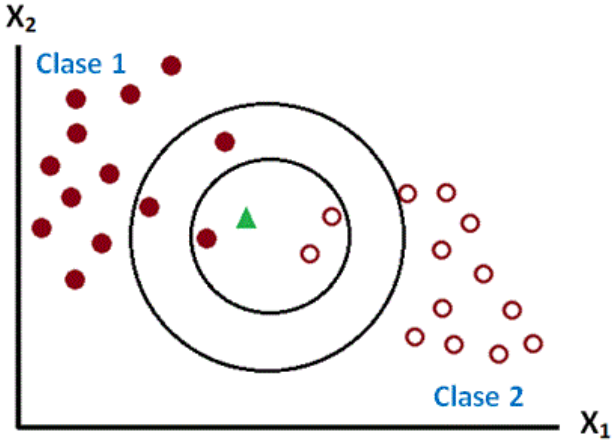

Figura 2.23: *Ejemplo de aplicación del algoritmo K-NN*

2.6.5. Etapa de control

Por último, la etapa de control o aplicación se corresponde con la que directamente interactúa con el usuario. Una vez que han sido detectadas las características y se han clasificado (o traducido) éstas en señales de control, la aplicación implementada debería realizar las acciones correspondientes. Que una aplicación pueda ser utilizada en la práctica depende de factores como la velocidad o la precisión. Algunas de estas aplicaciones permiten al usuario controlar dispositivos domóticos presentes en una vivienda, responder a preguntas del tipo SI/NO, manejar una silla de ruedas, movimiento de prótesis, etc.

2.7. Retos y futuro de BCI

Los principales problemas que reducen la confiabilidad y la exactitud de los BCI y que evita que esta tecnología sea clínicamente útil, son los problemas de interface sensorial y los problemas de traducción de algoritmos [18]. Con el fin de hacer clínicamente útil un BCI, la exactitud de la detección de la intención necesita ser muy alta y de hecho más alta que la actualmente

obtenida con diferentes tipos de BCI [18].

El primer obstáculo más importante para diseñar un BCI robusto es el problema de la interface sensorial. Hasta la fecha no existe una modalidad sensorial que sea adecuada y segura. Tres tipos de interfaces sensoriales han sido usados previamente, los sensores de EEG, ECoG y los microelectrodos implantados. Hay un punto medio entre seguridad y exactitud de los sensores.

De manera no sorpresiva, el BCI más exacto ha sido el basado en el implante cortical de micro-electrodos, el cual es el tipo de técnica más invasiva. Los microelectrodos están en contacto directo con las neuronas, lo que hace que relación señal ruido sea mejor que en otros tipos de sensores. Nicolelis demostró con qué exactitud los monos pueden ser entrenados para agarrar objetos usando un brazo robótico basado en BCI puesto en marcha directamente por las neuronas del mono encargadas de dicho movimiento [85]. Las mayores desventajas de este tipo de sensor son su alta invasividad en el tejido cerebral y su inestabilidad. La señal gradualmente se deteriorara debido al mecanismo defensivo del tejido cerebral que reacciona a objetos extraños. Esto causa la formación de tejido fibroso, lo que hace a los sensores inservibles luego de 6 meses [47].

El otro tipo de BCI es basado en EEG. El sensor de EEG tiene la mayor seguridad y la menor invasividad pero tiene la peor precisión y exactitud para acceder a la señal cerebral. A pesar de amplias investigaciones en este tipo de BCI durante los últimos 15 años, la exactitud del BCI basado en EEG no es promisoria para ayudar a pacientes a un control artificial protésico. No obstante, este tipo de BCI ha mostrado ser útil en funciones tan simples como *speller* (deletreador), el cual proporciona a los pacientes una lenta pero efectiva forma de comunicación [18].

La técnica intermedia entre precisión y seguridad es el BCI basado en ECoG, el cual ha demostrado ser promisorio. La matriz de sensores de electrodos son menos invasivas y proporcionan comparable precisión y resolución espacial comparado con el tipo de implante. El BCI basado en ECoG necesita mucho menos entrenamiento que el BCI basado en EEG y los investigadores han mostrado que puede ser lograda alta precisión y rápida repuesta en experimentos hechos en pacientes epilépticos que necesitan un implante ECoG temporal para identificar la localización exacta del foco epiléptico antes de la cirugía.

La comunidad BCI necesita un avance tecnológico en el campo de los sensores, que le permita abrir la puerta a una amplia área de aplicaciones que actualmente tienen gran impedimento debido a sus limitaciones por las actuales modalidades de sensores. Para proporcionar seguridad, no invasividad y acceso preciso a las señales cerebrales se requiere una tecnología de sensores diferente.

Es conocido que fMRI puede proporcionar un buen método no invasivo con una relativa alta resolución espacial. El problema con la técnica es su baja respuesta temporal. Las acciones neuronales pueden ser detectadas por los previos métodos en solo unos pocos milisegundos, mientras que la fMRI lo hace en unos pocos segundos. El otro problema con la técnica de fMRI y como se mencionó anteriormente es su costo y gran tamaño.

Recientemente un investigador de los laboratorios de Investigación de Telecomunicaciones Avanzadas en Neurociencia Computacional en Kyoto Japon, demostró la habilidad para reconstruir imágenes de $10 * 10$ pixeles de la corteza visual de un individuo mirando una imagen en blanco y negro con fMRI [76]. Los investigadores dicen que en futuras investigaciones se puede alcanzar una reconstrucción de una foto a color o incluso un sueño grabado en un archivo de video. Sin embargo, este logro está lejos de ser exitoso con la tecnología EEG, debido a la distorsión del sensor de EEG de las señales cerebrales.

fMRI portátil podría proporcionar una esperanza para la tecnología BCI. fMRI barato, portátil y preciso, podría ser un buen candidato para un BCI robusto. Investigadores del Laboratorio Nacional Lawrence Berkeley del departamento de Energía de EE. UU. han desarrollado exitosamente un laser basado en MIR portátil, barato y altamente sensible usando Magnetometría Atómica con mejor resolución temporal [119]. No obstante, para proporcionar un BCI rápido que sea capaz de controlar una prótesis artificial, esta nueva tecnología necesita desarrollarse más para incrementar su resolución temporal.

El otro reto radica en el cerebro en sí mismo. El cerebro usualmente tiene numerosos centros neuronales que cooperan para producir un simple movimiento de labios. La pregunta es si el cerebro es capaz de entrenar a específicas neuronas para que actúen como centros de redes de trabajo que controlen las extremidades a través de la vía de la espina dorsal produciendo un fino movimiento de la extremidad. La capacidad y la eficiencia del BCI dependen de la respuesta a esta pregunta [18].

Mejores algoritmos de traducción probablemente emergerán en el futuro, haciendo el control más realístico y más eficiente, sin embargo, la severidad y la naturaleza de este déficit en las actuales investigaciones de BCI indican que debería ser desarrollada una traducción de algoritmos más realista que reduzca esta dificultad.

Parte III

Datos y su caracterización

Capítulo 3

Adquisición de datos y su caracterización

Para el desarrollo del TFM se emplea un conjunto de datos de EEG reales, proporcionado por una competición a nivel mundial "BCI Competitions II" de 2003 [2], propuesto por el grupo *BerlinBrain-Computer Interface* en el que diferentes grupos de investigación trataban de decodificar señales EEG accesibles vía web. El objetivo de la competición es validar modelos de procesado de señal y métodos de clasificación que puedan servir para su aplicación a problemas de BCI.

Desde el año 2001, se ha llevado a cabo cada dos años una serie de competiciones BCI en las que cualquier investigador puede demostrar la eficiencia de su sistema BCI empleando una serie de señales cerebrales proporcionadas por algunos de los más importantes grupos de investigación sobre BCI.

El objetivo principal de las competiciones BCI es impulsar la investigación en BCI. Se ofrecen diferentes tipos de señales. Las competiciones proporcionan un conjunto de señales que pueden ser utilizadas por todos los investigadores de la comunidad científica. Esto hace más sencilla la comparación de los diferentes métodos presentados por los investigadores facilitando la replicación de los estudios.

En este capítulo, se detalla los conjuntos de datos "*Ia*" y "*Ib*" de "BCI Competitions II"[2] que serán los empleados para realizar el estudio de los métodos de extracción y clasificación de características. También, se resume los métodos aplicados por el ganador de "BCI Competitions II"[2] para el procesamiento de las señales en ambos conjuntos de datos. Finalmente, se hace un primer análisis visual y estadístico de dichos conjuntos de datos.

3.1. Competición BCI II

El conjunto de datos "*I*" de "BCI Competitions II" [2] ha sido proporcionado por el Instituto de Psicología Médica y Neurobiología Conductual de la Universidad de Tübingen. Este conjunto de datos, contiene dos experimentos diferentes denominados "*Ia*" y "*Ib*" que son datos de señales de EEG relacionadas con la autorregulación cortical. Este tipo de señales son conocidas como Potenciales Corticales Lentos.

3.1.1. Descripción del conjunto de datos "*Ia*"

Los datos de EEG del conjunto de datos "*Ia*" son tomados de un sujeto sano. Al sujeto se le pide que piense en mover un cursor hacia arriba y hacia abajo de una pantalla del computador (dos tareas mentales), mientras se registran sus potenciales corticales.

Durante la grabación el sujeto recibe *feedback* visual de sus SCP (Cz-mastoides). La positividad cortical (que está relacionada con la reducción de de la activación cortical) conduce a un movimiento del cursor hacia abajo de la pantalla y la negatividad cortical (que está relacionada con la activación cortical, asociada a la realización de movimientos) conduce a un movimiento del cursor hacia arriba. La Figura 3.1 muestra un ejemplo aclarativo sobre positividad y negatividad cortical.

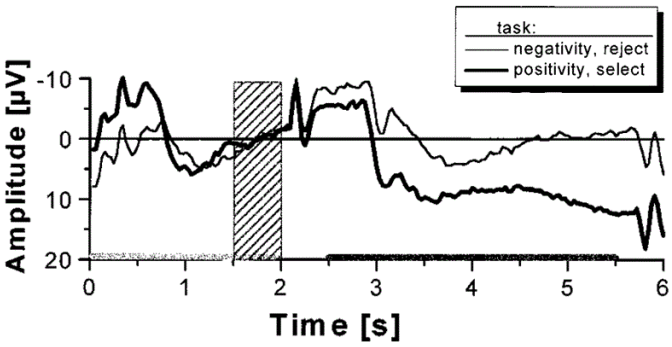

Figura 3.1: *Registro EEG de un BCI basado en SCP. SCP promedio de un paciente durante 700 ensayos para las fases de preparación (izquierda del rectángulo) toma del umbral de comparación (rectángulo) y fase activa (derecha del rectángulo). Durante la fase activa, la línea gruesa indica una elección binaria (positividad cortical) mientras que la línea fina indica la opuesta (negatividad cortical). Fuente: [25].*

Cada ensayo dura 6 s. Durante cada ensayo, se presenta visualmente una tarea (rectángulo verde) resaltada arriba o abajo de la pantalla para indicar negatividad o positividad desde los 0,5 s hasta el final de la prueba. El *feedback* visual es presentado desde el segundo 2 hasta el segundo 5,5. Sólo el intervalo de 3,5 s de cada ensayo es proporcionado para el *Training* y el *Testing*. La Figura 3.2 muestra el paradigma del experimento.

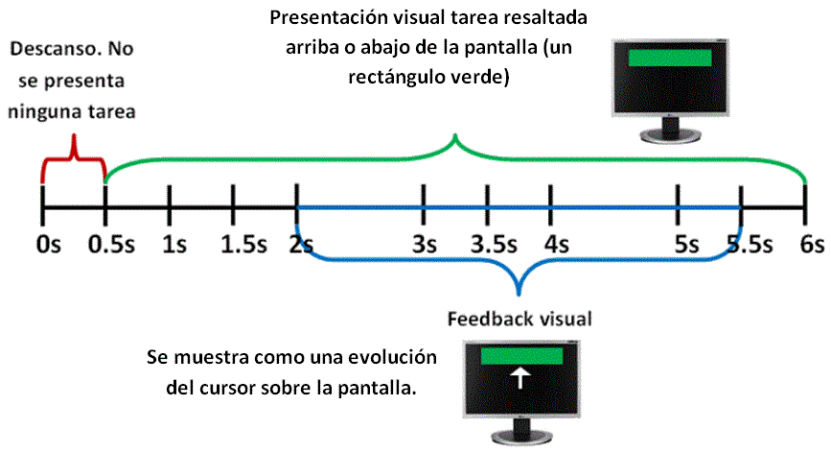

Figura 3.2: *Experimento del conjunto de datos "Ia" de la Competición BCI II.*

El *feedback* visual está representado por un cursor que aparece entre el intervalo 2,5 s a 5,5 s en la parte central de la pantalla y que se mueve hacia el objetivo o en dirección contraria del mismo, lo que le permitirá al sujeto saber si está realizando correctamente la tarea mental (positividad cortical mueve cursor hacia abajo y negatividad cortical mueve cursor hacia arriba).

Se registran seis canales de datos de EEG (denotación que sigue el sistema 10/20), como se muestra en la Figura 3.3.

Canal 1: A1-Cz (A1 = mastoides izquierda)
Canal 2: A2-Cz (A2 = mastoides derecha)
Canal 3: (2 cm frontal de C3)-Cz = F3-Cz
Canal 4: (2 cm parietal de C3)-Cz = P3-Cz
Canal 5: (2 cm frontal de C4)-Cz = F4-Cz
Canal 6: (2 cm parietal de C4)-Cz = P4-Cz

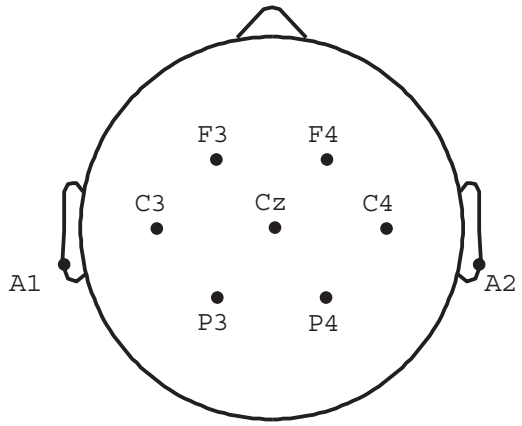

Figura 3.3: *Mapa de electrodos de los datos de EEG de estudio para 6 canales. Los 6 canales están formados por A1,A2,F3,P3,F4 y P4 referenciados a Cz.*

La frecuencia de muestreo es de 256 Hz y la duración de la grabación de 3,5 s que generan 896 muestras por canal para cada ensayo. Como se usaron 6 canales, la señal total se compone de 5376 muestras.

Se registran en total 268 ensayos en dos días diferentes y se mezclan aleatoriamente. Del total global de ensayos, 168 son grabados en el primer día y los 100 ensayos restantes en el segundo día. Las matrices que conforman el conjunto de datos de *Training*, contienen datos de 135 ensayos pertenecientes a la "clase 0" y 133 ensayos pertenecientes a la "clase 1". Las dimensiones de las matrices son $135 * 5377$ y $133 * 5377$ respectivamente. Cada línea de una matriz contiene los datos de un ensayo. Los primeros códigos de las columnas son las etiquetas de referencia para cada una de las clases (0/1). Las columnas restantes contienen las muestras de tiempo de los 6 canales de EEG que comienza con 896 muestras para el canal 1 y termina con 896 muestras para el canal 6.

La matriz que conforma el conjunto de datos de *Testing* tiene una dimensión de $293 * 5376$ y contiene 293 ensayos de datos de prueba. Cada ensayo se registra en el segundo día y pertenece tanto a la "clase 0" como a la "clase 1". La matriz de *Testing* tiene casi la misma estructura que las matrices de datos de *Training* con la diferencia de que falta la etiqueta de las dos respectivas clases. Cada línea contiene 896 muestras de tiempo de cada uno de los 6 canales de EEG.

3.1.2. Descripción conjunto de datos "*Ib*"

Los datos de EEG del conjunto de datos "*Ib*" son tomados de un paciente con ALS con ventilación artificial. Al sujeto se le pide que piense en mover un cursor hacia arriba y hacia abajo de una pantalla del computador (dos tareas mentales), mientras se registran sus potenciales corticales.

Durante la grabación el sujeto recibe *feedback* visual y auditivo de sus SCP (Cz-mastoides). La positividad cortical (que está relacionada con la reducción de de la activación cortical) conduce a un movimiento del cursor hacia abajo de la pantalla y la negatividad cortical (que está relacionada con la activación cortical, asociada a la realización de movimientos) conduce a un movimiento del cursor hacia arriba. La Figura 3.1 muestra un ejemplo aclarativo sobre positividad y negatividad cortical.

Cada ensayo dura 8 s. Durante cada ensayo, se presenta visual y auditivamente una tarea (rectángulo verde) resaltada en la parte superior de la pantalla (para negatividad) o en la parte inferior de la pantalla (para positividad) desde el segundo 0,5 hasta los 7,5 s de cada ensayo. Adicionalmente la tarea ("arriba" o "abajo") es vocalizada a los 0,5 s. El *feedback* visual es presentado desde el segundo 2 a los 6,5 segundos. Solamente este intervalo de 4,5 s de cada ensayo es proporcionado para el *Training* y el *Testing*. La Figura 3.4 muestra el paradigma del experimento.

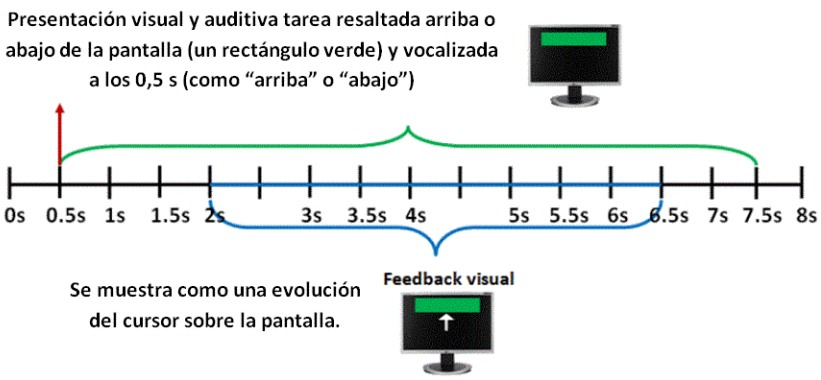

Figura 3.4: *Experimento del conjunto de datos "Ib" de la Competición BCI II.*

Además de los mismos seis canales usados en el conjunto de datos "Ia", hay otro canal para detectar los artefactos de los movimientos oculares verticales (vEOG).

Canal 1: A1-Cz (A1 = mastoides izquierda)
Canal 2: A2-Cz (A2 = mastoides derecha)
Canal 3: (2 cm frontal de C3)-Cz = F3-Cz
Canal 4: (2 cm parietal de C3)-Cz = P3-Cz
Canal 5: vEOG
Canal 6: (2 cm frontal de C4)-Cz = F4-Cz
Canal 7: (2 cm parietal de C4)-Cz = P4-Cz

La frecuencia de muestreo es de 256 Hz y la duración de la grabación de 4,5 s que genera 1152 muestras por canal para cada ensayo. Como se usaron 7 canales, la señal total se compone de 8064 muestras.

Los ensayos se registran en el mismo día y se mezclan aleatoriamente. Las matrices que conforman el conjunto de datos de *Training*, contienen datos de 100 ensayos pertenecientes a la "clase 0" y 100 ensayos pertenecientes a la "clase 1". Las dimensiones de las matrices son $100 * 8065$ y $100 * 8065$ respectivamente. Cada línea de una matriz contiene los datos de un ensayo. Los primeros códigos de las columnas son las etiquetas de referencia para cada una de las clases (0/1). Las columnas restantes contienen las muestras de tiempo de los 7 canales de EEG/EOG que comienza con 1152 muestras para el canal 1 y finaliza con 1152 muestras para el canal 7.

La matriz que conforma el conjunto de datos de *Testing* tiene una dimensión de $180 * 8064$ y contiene 180 ensayos de datos de prueba. Estos datos de prueba se registran después de los datos de *Training* (en el mismo día). La matriz de *Testing* tiene casi la misma estructura que las matrices de datos de *Training* con la diferencia de que falta la etiqueta de las clases 0 y 1. Cada línea contiene 1152 muestras de tiempo de cada uno de los 7 canales de de EEG/EOG.

3.2. Descripción del método ganador

3.2.1. Método ganador del conjunto de datos "Ia"

El método ganador de "BCI Competitions II" [2] para el conjunto de datos "Ia" fue el desarrollado por Brett Mensh del Instituto de Tecnología de

Massachussets en Cambridge, obteniendo un porcentaje de error en la clasificación *online* de 11,3 % [11], [73] .

Para el conjunto de datos "*Ia*" el autor calcula la Componente Continua (*Direct Current*, DC) de los últimos dos tercios o menos de las señales para cada canal. Esto proporciona una dimensión para cada uno de los canales 1 y 2. Luego hace un análisis espectral utilizando el método Multitaper de Thomson (*Thomson Multi Taper Method*, MTM), y los canales 4 y 6 muestran potencias diferentes en el rango de frecuencias de 24-37 Hz entre las dos condiciones ("arriba" o "abajo"). MTM es un método de estimación espectral y reconstrucción de la señal temporal que contiene tanto componentes continuas y singulares [105].

Utiliza las características proporcionadas por los canales 1,2,4 y 6 para hacer la clasificación con LDA.

Para el conjunto de datos "*Ib*" no utiliza el nivel DC debido a la alta complejidad de las señales. En primer lugar, emplea un filtro espacial Laplaciano, que tiene el efecto de enfatizar las diferencias entre canales individuales por encima de efectos globales mayores. Las características de las señales son extraídas mediante un análisis espectral de potencia de la banda beta. Los resultados de la clasificación para éste conjunto de datos no son publicados.

3.2.2. Método ganador del conjunto de datos "Ib"

El método ganador de "BCI Competitions II" [2] para el conjunto de datos "*Ib*" fue el desarrollado por Vladimir Bostanov de la Universidad de Tübingen, obteniendo un porcentaje de error en la clasificación *online* de 45,6 % [12], [26].

Utiliza la Trasnformada Wavelet Continua (con un escalograma para detectar los picos) para extraer las características de las señales cerebrales y clasificarlas con LDA. Para obtener una presentación más concisa de la señal, utiliza un filtro paso baja que elimina las frecuencias más altas y no tiene en cuenta el canal 5 en el procesamiento de los datos. En [26] se encuentra los detalles más específicos acerca del procesamiento de las señales de SCP para éste conjunto de datos.

3.3. Visualización y caracterización

3.3.1. Caracterización del conjunto de datos "Ia"

Con el propósito de identificar las características de los datos de SCP para el conjunto de datos "*Ia*" que puedan discriminar entre las dos clases de referencia para cada ensayo (0/1), se separan los 268 ensayos que conforman el conjunto de *Training*, en los ensayos que están referenciados con la etiqueta 0 de los que están referenciados con la etiqueta 1.

Para cada uno de los seis canales de EEG utilizados, se calcula y se representa gráficamente el promedio o la media de todos los ensayos de cada una de las clases (135 ensayos pertenecientes a la "clase 0" y 133 ensayos pertenecientes a la "clase 1") en el dominio del tiempo, como se puede ver en las Figuras 3.5, 3.7, 3.9, 3.11, 3.13 y 3.15.

Se hace también una representación gráfica de cada canal utilizando un histograma para obtener una "primera vista" general de la distribución de las muestras (nivel DC) respecto a cada una de las clases. Dicho procedimiento se ve reflejado en las Figuras 3.6, 3.8, 3.10, 3.12, 3.14 y 3.16.

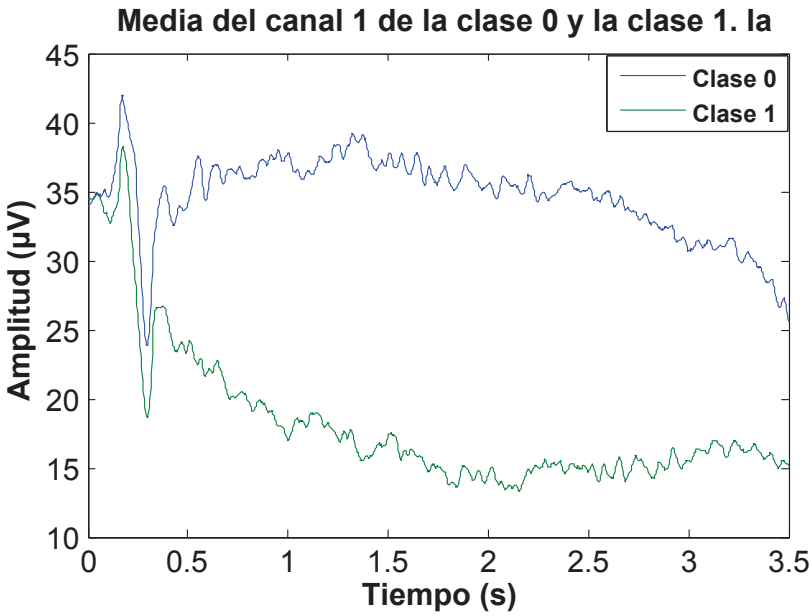

Figura 3.5: *Representación de la media de los ensayos del canal 1 de la clase 0 y de la clase 1 para el conjunto de datos "Ia".*

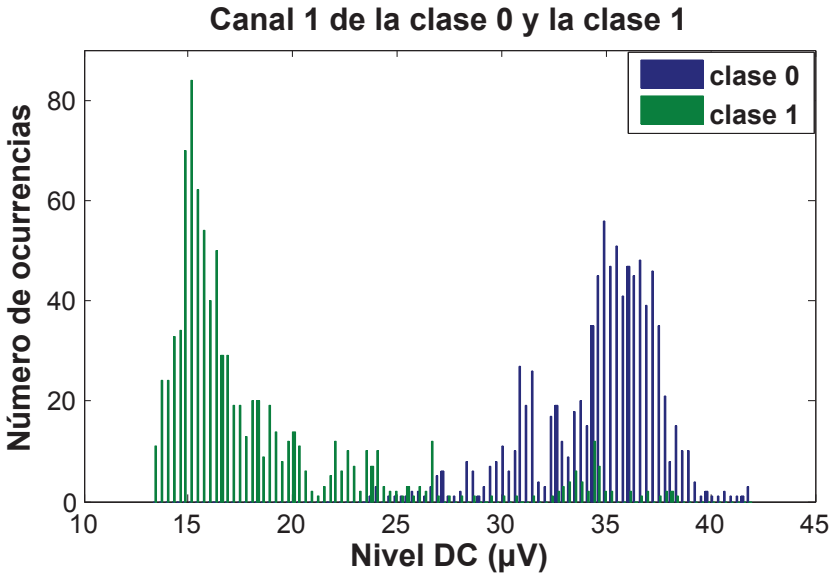

Figura 3.6: *Histograma del canal 1 de la clase 0 y de la clase 1 para el conjunto de datos "Ia".*

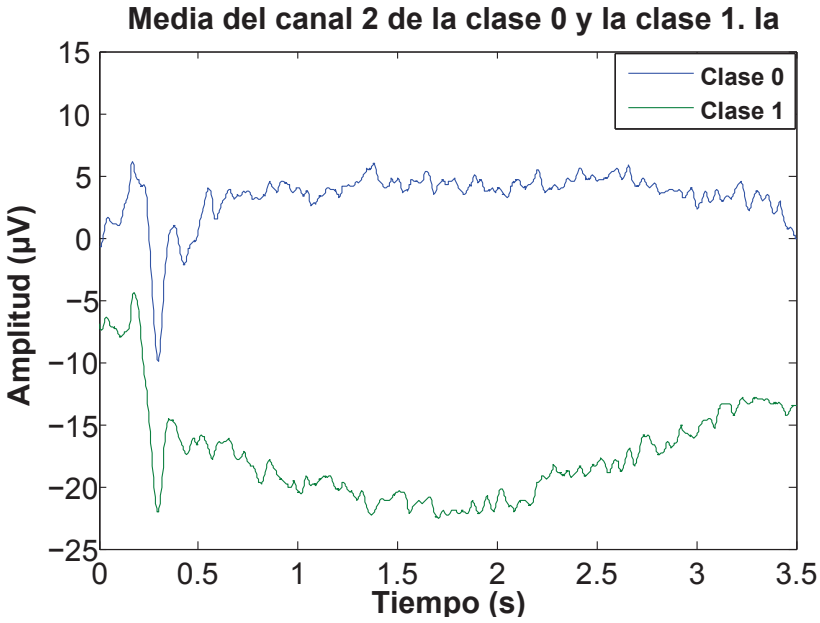

Figura 3.7: *Representación de la media de los ensayos del canal 2 de la clase 0 y de la clase 1 para el conjunto de datos "Ia".*

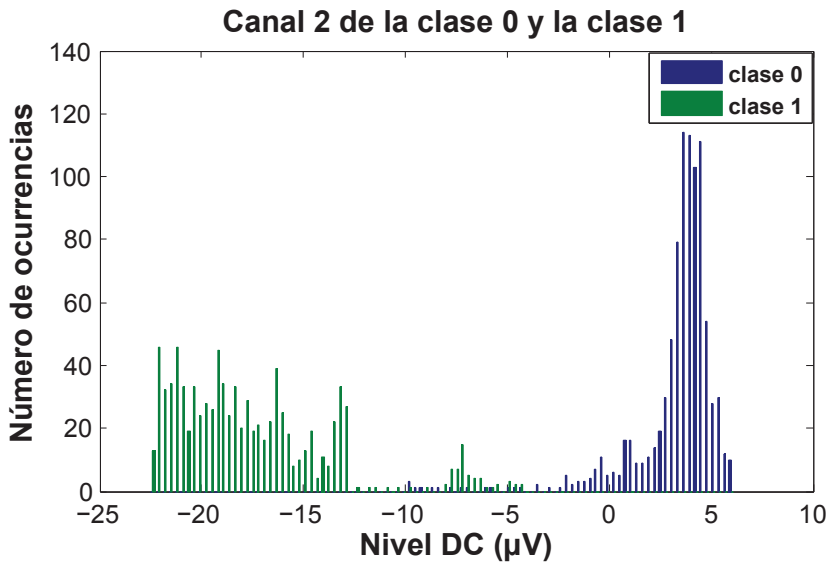

Figura 3.8: *Histograma del canal 2 de la clase 0 y de la clase 1 para el conjunto de datos "Ia".*

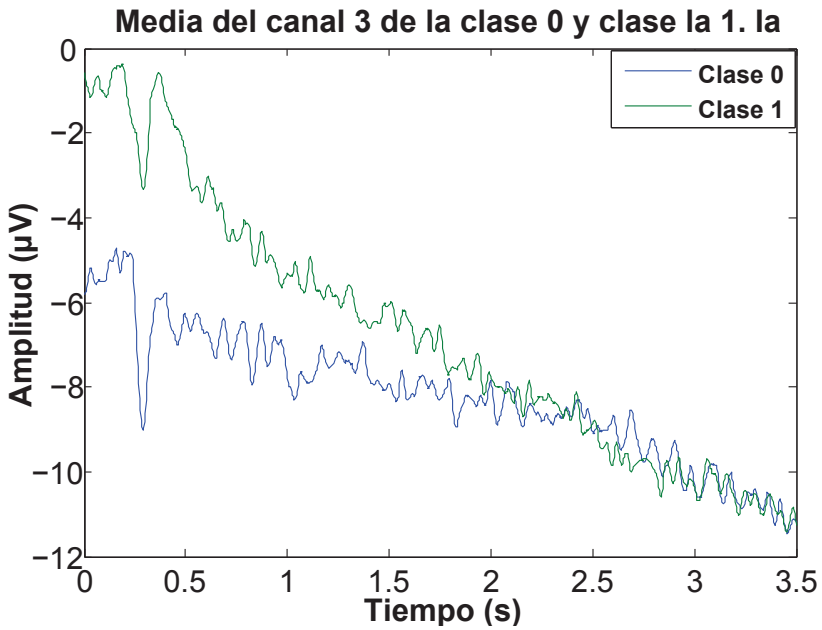

Figura 3.9: *Representación de la media de los ensayos del canal 3 de la clase 0 y de la clase 1 para el conjunto de datos "Ia".*

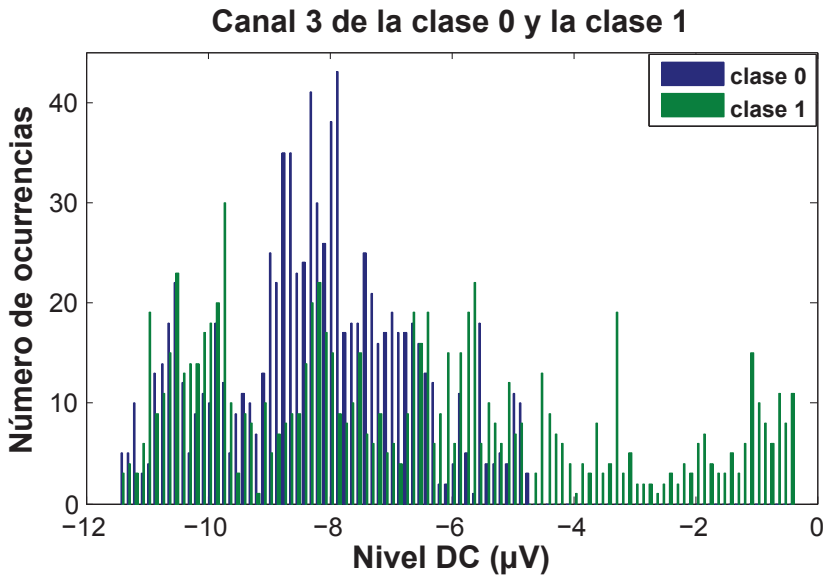

Figura 3.10: *Histograma del canal 3 de la clase 0 y de la clase 1 para el conjunto de datos "Ia".*

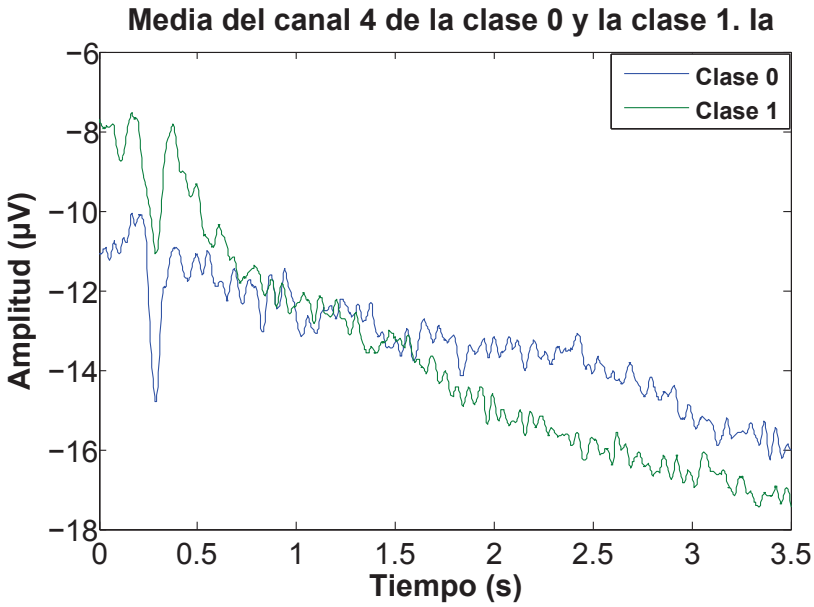

Figura 3.11: *Representación de la media de los ensayos del canal 4 de la clase 0 y de la clase 1 para el conjunto de datos "Ia".*

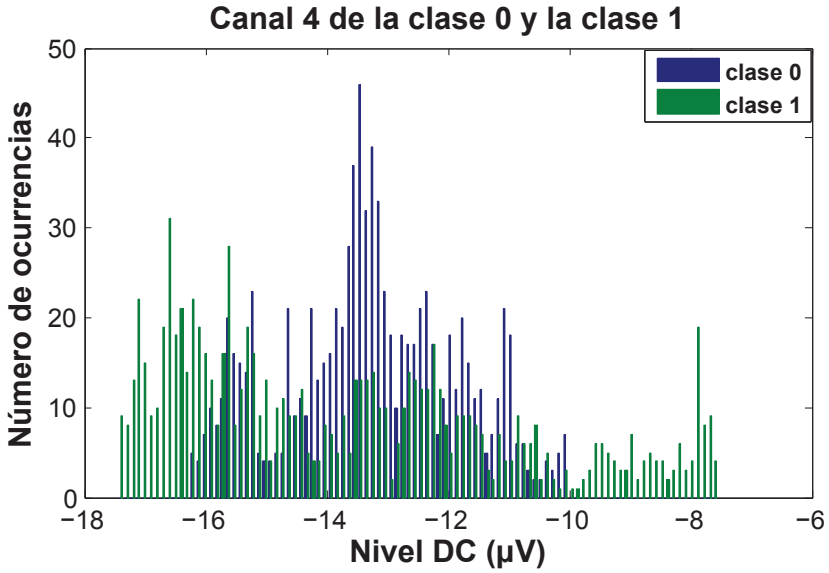

Figura 3.12: *Histograma del canal 4 de la clase 0 y de la clase 1 para el conjunto de datos "Ia".*

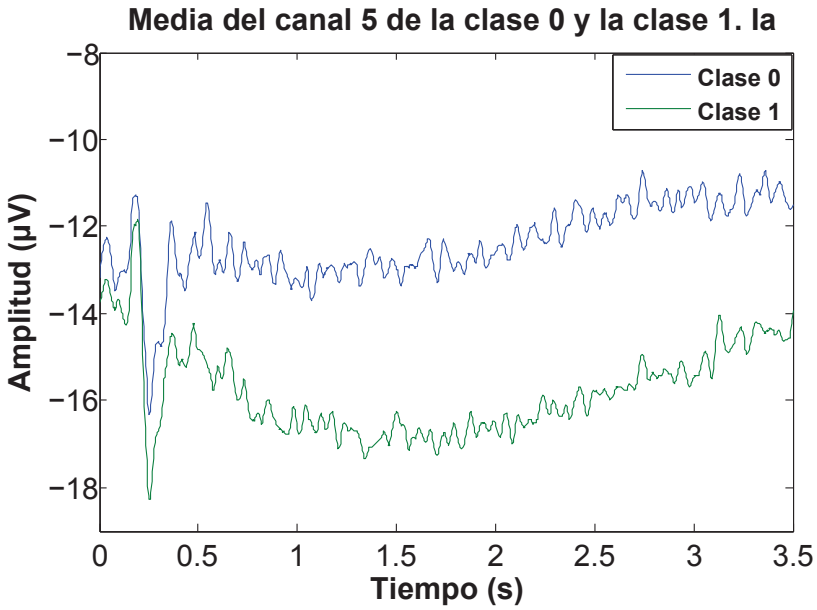

Figura 3.13: *Representación de la media de los ensayos del canal 5 de la clase 0 y de la clase 1 para el conjunto de datos "Ia".*

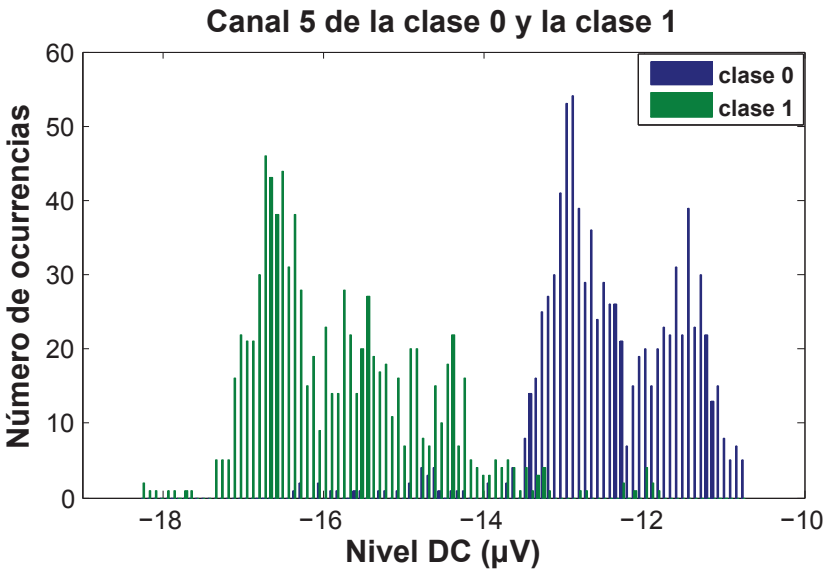

Figura 3.14: *Histograma del canal 5 de la clase 0 y de la clase 1 para el conjunto de datos "Ia".*

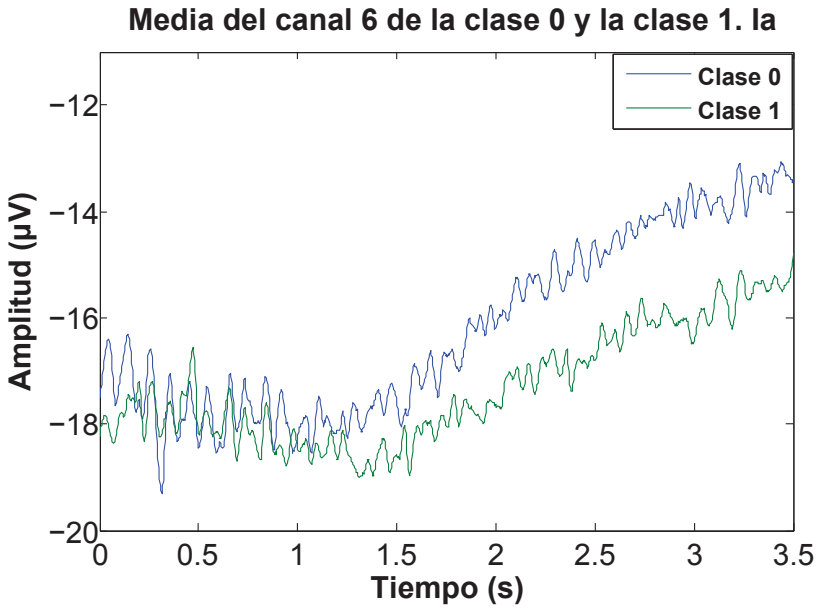

Figura 3.15: *Representación de la media de los ensayos del canal 6 de la clase 0 y de la clase 1 para el conjunto de datos "Ia".*

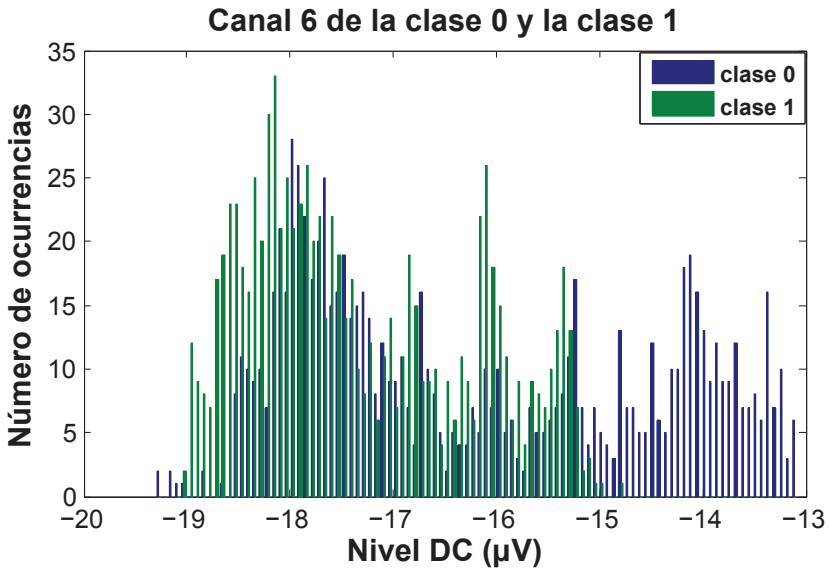

Figura 3.16: *Histograma del canal 6 de la clase 0 y de la clase 1 para el conjunto de datos "Ia".*

Después de analizar todas y cada una de las Figuras anteriormente representadas que corresponden al conjunto de datos "Ia", se puede ver en las Figuras 3.5 y 3.7 una clara diferencia de separación entre la "clase 0" y la "clase 1".

Así mismo, las Figuras 3.6 y 3.8 describen el comportamiento de los datos de SCP de los canales 1 y 2, mostrando una gran diferencia en la dispersión y separación de las muestras pertenecientes a cada una de las clases.

Los canales 1 y 2 son los más precisos en proyectar dicha separación entre las clases, y por ello, las señales de estos canales, serán utilizados para llevar a cabo el procesamiento inicial con diferentes técnicas estadísticas, como el nivel DC y el valor de la Varianza, que permitan extraer sus características.

A diferencia del método ganador de "BCI Competitions II" [2] en el conjunto de datos "Ia" [11], nosotros calculamos en nivel DC para diferentes segmentos en cada ensayo procurando con esto, seguir la dinámica no estacionaria de las señales de SCP. El objetivo de segmentar las señales y obtener el nivel DC de los mismos, es conseguir una representación simbólica de la serie temporal que permita con ello, la reducción de la dimensionalidad sin perder la correlación con la serie de tiempo original [68].

Consideramos que podría ser útil, obtener los valores aproximados bajo la curva, calculados por medio de una Integral, usando 3 diferentes combinaciones de canales: (1) canales 1 y 2 (2) canales 1, 2 y 5 y (3) todos los 6 canales. Estos valores podrían aportar información significativa considerando la diferencia clara entre las clases de los canales 1, 2 y seguramente el canal 5, ya que como puede verse en las Figuras 3.13 y 3.14 que representan la media y la distribución de las muestras de la "clase 0" y la "clase 1", es el tercer mejor canal en el que se visualiza una pequeña diferencia entre las clases y además el histograma ilustra una diferenciación entre la dispersión de las muestras de cada una de las clases a considerar. Al utilizar todos los canales tendríamos más datos y posiblemente mayor información significativa.

En el resto de canales difícilmente se puede ver la diferenciación entre las clases, posiblemente debido al ruido derivado de la medición de SCP. Por ello, se emplearán métodos de extracción de características en el dominio de la frecuencia y/o tiempo-frecuencia.

3.3.2. Caracterización del conjunto de datos "Ib"

Para identificar las características de los datos de SCP en el conjunto de datos "*Ib*" que puedan discriminar entre la "clase 0" y la "clase 1", se separa los 200 ensayos de *Training*, en aquellos referenciados con la etiqueta 0 y la etiqueta 1 (como se hizo en *"Ia"*).

Para cada uno de los 7 canales de EEG/EOG utilizados, se calcula y se representa gráficamente el promedio o la media de todos los ensayos de cada una de las clases en el dominio del tiempo, como se puede ver en las Figuras 3.17, 3.18, 3.19, 3.20, 3.21, 3.22 y 3.23 .

Los canales 4, 5 y 6 representados en las Figuras 3.20, 3.21 y 3.22 son los que quizás visualmente tengan una mínima diferencia de separación entre las clases. Por ello, se analizará dicha combinación de canales con las técnicas estadísticas propuestas: el nivel DC, la Varianza y la Integral.

Para el resto de canales, las señales resultantes son considerablemente más ruidosas. En general, las señales adquiridas del paciente con ALS para este conjunto de datos *"Ib"*, con malas, ya que a simple vista no se puede ver un patrón evidente inmerso.

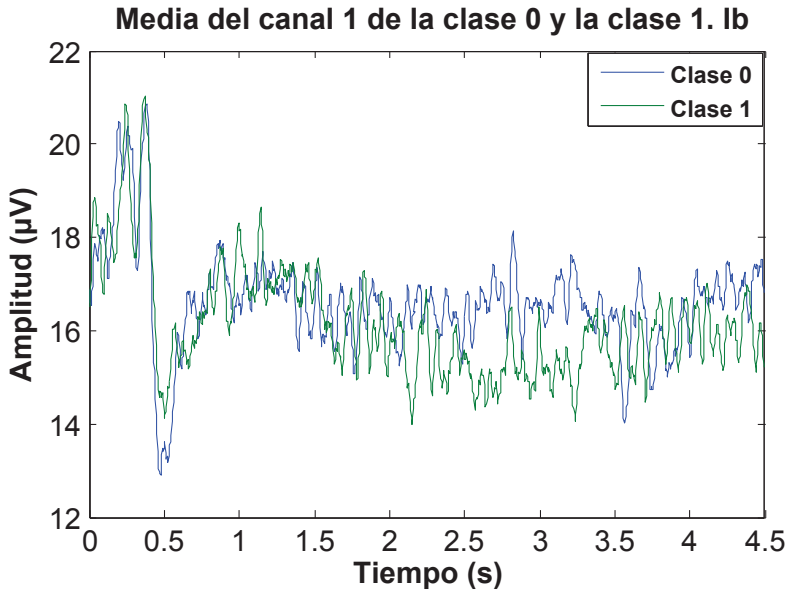

Figura 3.17: *Representación de la media de los ensayos del canal 1 de la clase 0 y de la clase 1 para el conjunto de datos "Ib".*

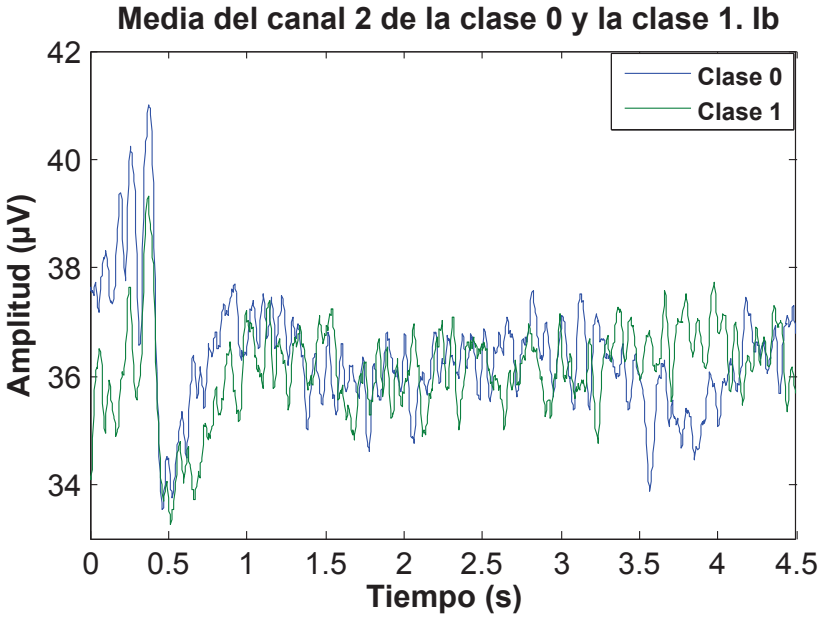

Figura 3.18: *Representación de la media de los ensayos del canal 2 de la clase 0 y de la clase 1 para el conjunto de datos "Ib".*

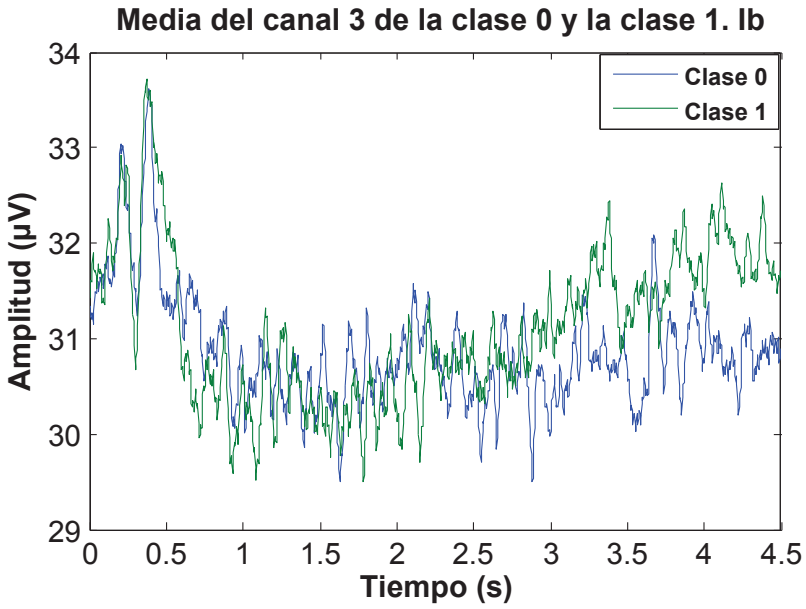

Figura 3.19: *Representación de la media de los ensayos del canal 3 de la clase 0 y de la clase 1 para el conjunto de datos "Ib".*

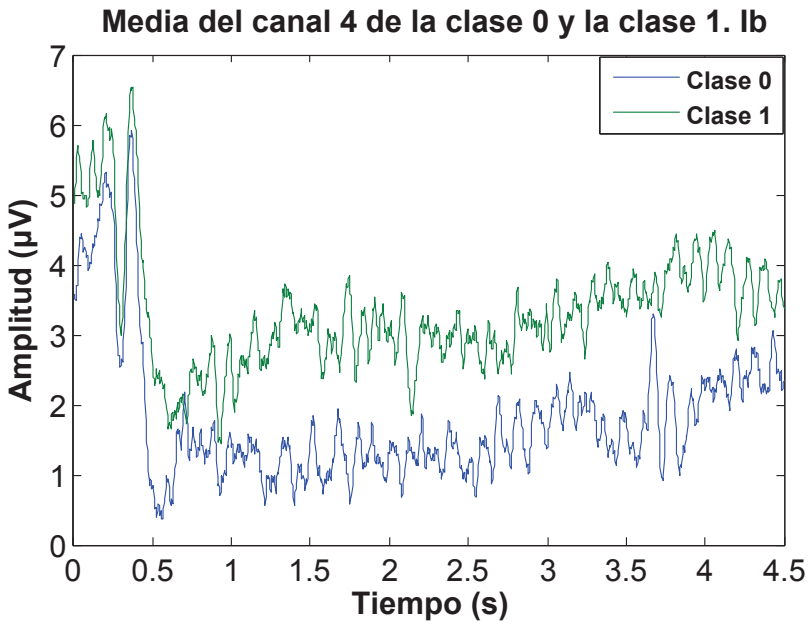

Figura 3.20: *Representación de la media de los ensayos del canal 4 de la clase 0 y de la clase 1 para el conjunto de datos "Ib".*

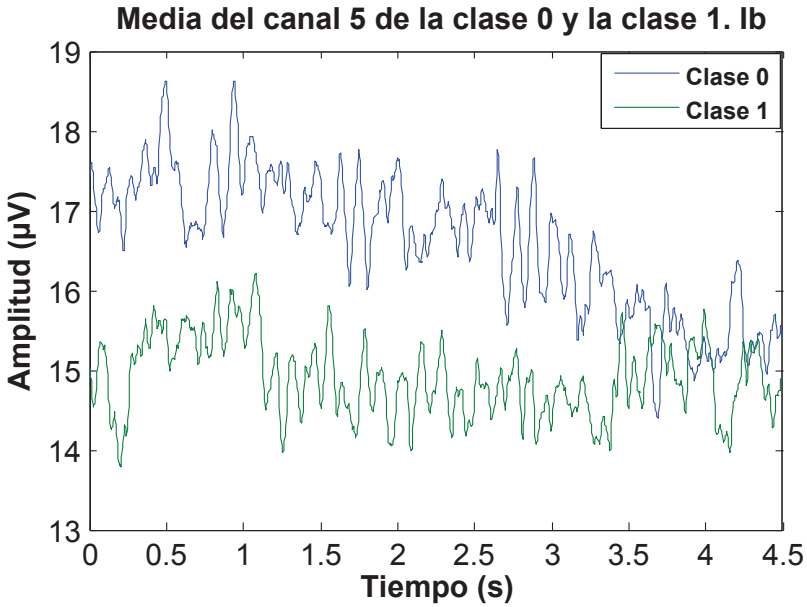

Figura 3.21: *Representación de la media de los ensayos del canal 5 de la clase 0 y de la clase 1 para el conjunto de datos "Ib".*

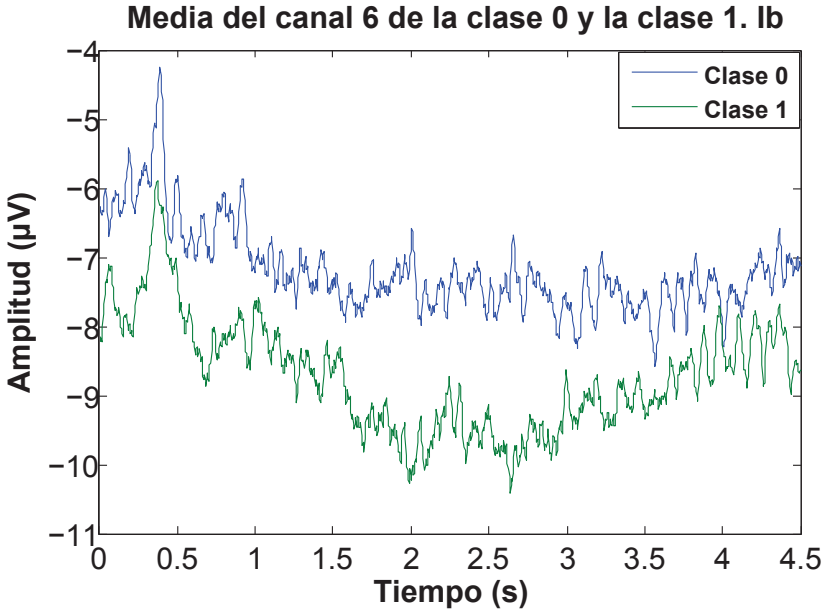

Figura 3.22: *Representación de la media de los ensayos del canal 6 de la clase 0 y de la clase 1 para el conjunto de datos "Ib".*

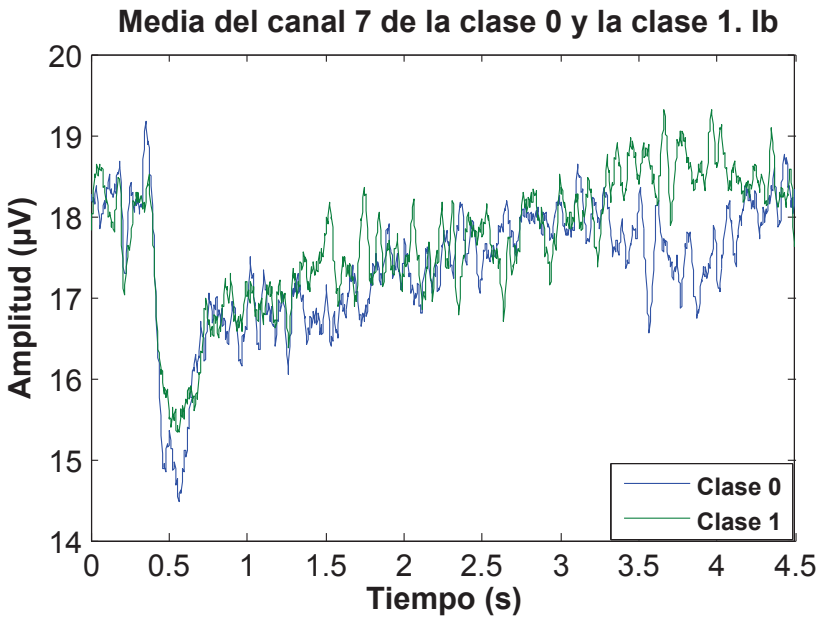

Figura 3.23: *Representación de la media de los ensayos del canal 7 de la clase 0 y de la clase 1 para el conjunto de datos "Ib".*

Emplear las características obtenidas con las 3 técnicas estadísticas planteadas no sería suficiente para caracterizar las diferencias entre positividad y negatividad cortical en este conjunto de datos. Sin embargo, se podrían filtrar las señales y quedarnos solamente con las bajas frecuencias y omitir el canal 5 que ha sido utilizado para detectar los artefactos de los movimientos oculares verticales (vEOG), tal como lo propuso, el ganador de "BCI Competitions II" [26]. Para el resto de canales, es necesario utilizar otros métodos de extracción de características ya sea en el dominio de la frecuencia y/o tiempo-frecuencia.

Con el objetivo de obtener vectores con características significativas que nos permitan caracterizar las diferencias entre positividad y negatividad cortical para los conjuntos de datos *"Ia"* y *"Ib"*, proponemos crear vectores heterogéneos o súper vectores, que contengan características extraídas con los mejores métodos y técnicas de extracción.

Los resultados de la clasificación en los escenarios *offline* y *online simulado* para ambos conjuntos de datos de SCP con el vector de características formado por el Nivel DC, el valor de la medida de la Varianza y los valores de aproximación del área bajo la curva, se muestran en el Capítulo 5, sección 5.2.

Parte IV

Estudio de métodos de extracción y clasificación

Capítulo 4

Aplicación de métodos

El procesado de la señal en sistemas BCI se divide habitualmente en cinco etapas, como se ha explicado en secciones anteriores. En primer lugar la adquisición de la señal, luego la etapa de pre-procesado que prepara las señales adquiridas en la etapa anterior en una forma adecuada para su posterior procesamiento. La tercera etapa que consiste en la extracción de determinadas características específicas de la señal EEG que pueden ser útiles para discriminar entre diferentes tareas mentales. A continuación, se aplican métodos de selección de características que escogen las más significativas dentro del conjunto extraído, que codifican la intención del usuario. Finalmente, los algoritmos de clasificación traducen el conjunto de características seleccionado en un comando concreto, relacionado con la intención del usuario.

En este capítulo se presenta en detalle la aplicación de los métodos utilizados en la etapa de extracción y clasificación de características en los conjuntos de datos "Ia" y "Ib" de SCP en el escenario *offline* y su extrapolación en el escenario *online simulado*. Los métodos de extracción de características que incluye este TFM son PCA, modelos AR y DWT. Finalmente, se utiliza LDA y una SVM para hacer la clasificación de las características.

4.1. Escenarios de análisis de datos: *offline* y *online simulado*

La aplicación de los métodos de extracción y clasificación de características empleados en este trabajo, se ha realizado desde dos escenarios o enfoques diferentes: mediante un análisis *offline* de los datos y su extrapolación en el escenario *online simulado*.

El análisis *offline* nos permite estimar ciertas propiedades de las señales observándolas en su extensión en el tiempo y realizar cálculos que implican largos tiempos de computación. Las señales de estudio de SCP, se han descargado de [2] y almacenado posteriormente en el computador. En este tipo de análisis, se prueba un método en concreto que permita extraer las características de dichas señales para su posterior clasificación.

El análisis *online* es el medio para evaluar un BCI en un entorno real. Teniendo en cuenta, que en los sistemas BCI el análisis de las señales es llevado a cabo mediante un análisis *offline* y *online*, nuestra propuesta en este TFM, es hacer un análisis intermedio *online simulado*, con el propósito de obtener una estimación de los resultados como si el usuario hubiera utilizado el sistema en tiempo real, asegurando con ello, un mayor éxito en dicho entorno.

En el análisis *online simulado*, las señales son analizadas y procesadas por segmentos o trozos de un determinado tamaño en el tiempo (ventana). La ventana se va incrementando y desplazando a lo largo de toda la señal de EEG con el objetivo de determinar el segmento mínimo y óptimo que nos permita detectar los eventos (mover cursor arriba o abajo de una pantalla de computador) con el menor porcentaje de error.

4.2. Extracción de características

Se busca, crear una representación manejable y significativa de la señal de EEG, con vistas a maximizar el éxito potencial de la fase de clasificación y a su vez el rendimiento global del sistema. A continuación se presenta cada uno de los métodos utilizados para extraer las características de los datos de SCP.

4.2.1. Análisis de Componentes Principales

El Análisis de Componentes Principales (*Principal Component Analysis*, PCA) es un método clásico de análisis de datos que tiene sus principales aplicaciones en el campo de la extracción de características y compresión de datos.

PCA decorrelaciona las señales fuentes en EEG, transformando un set multivariado de datos con n componentes correlacionados $X = [x_1|x_2|...x_n]$, en un set de componentes no correlacionados llamados Componentes Principales (*Principal Components, PCs*), encontrando las direcciones ortogonales de

máxima varianza en las señales EEG.

La matriz de covarianza C; es calculada de los datos correlacionados de entrada. Los eigenvectores de C; son calculados y se crea una matriz modal $E = [e_1|e_2|...e_n]$. Los datos de entrada son transformados mediante:

$$Y = E^T X \tag{4.1}$$

La nueva matriz Y consta de las componentes principales concatenadas, en la Ecuación 4.2 la varianza de cada componente principal es encontrada a partir de la matriz de covarianza:

$$var(Y_i) = E_i^T C \qquad i = 1, 2, ...n \tag{4.2}$$

PCA rota el espacio de datos original de tal forma que los ejes del nuevo sistema de coordenadas apuntan en las direcciones más altas de varianza de los datos. Los ejes o nuevas variables son los PCs y se ordenan según su varianza, de tal manera que el primer componente *PC1* represente la dirección de la mayor varianza posible. La dirección del segundo componente *PC2*, representa las más alta varianza ortogonal restante del primer componente.

Esto puede ser naturalmente extendido para obtener el número requerido de componentes que juntos abarquen un espacio del componente cubriendo de esta forma la cantidad deseada de la varianza [64].

Ya que los componentes describen direcciones específicas en los espacios de datos, cada componente depende de cierta cantidad de cada una de las variables originales: cada componente es una combinación lineal de todas las variables originales.

A menudo se puede asumir que la varianza baja representa un ruido de fondo no deseado. La dimensionalidad de los datos puede sin embargo ser reducido, sin pérdida de información relevante, mediante la extracción del espacio del componente dimensional más bajo, cubriendo la varianza alta. Usar un menor número de componentes principales en vez de datos originales de alta dimensión, es un paso común de pre-procesamiento que a menudo mejora los resultados de los análisis subsecuentes tales como una clasificación.

En MatLab, el análisis de componentes principales hace parte del *toolbox* estadístico. Hemos utilizado la función ***princomp*** que realiza el PCA en X (siendo X una matriz que contiene los datos del conjunto de *Training* de "Ia", en este caso, con una dimensión de $268 * 5376$) y devuelve los PCs. Las filas de X corresponden a las observaciones y las columnas a las variables. La sintaxis de la función ***princomp*** es:

[coefs,scores,variances] = princomp (X)

donde *coefs*, es una matriz de $5376 * 5376$ en la que cada columna contiene los coeficientes de las combinaciones lineales de las variables originales que generan los PCs. Las columnas están en orden decreciente de varianza de los componentes.

La segunda salida corresponde a los *scores*, las puntuaciones de los PCs, es decir, las coordenadas de los datos originales (X) en el espacio definido por los PCs. Esta salida es del mismo tamaño que la matriz de datos de entrada (X). La matriz *scores*, contiene información en las primeras 5376 columnas que representan el valor de los datos proyectados en *PC1*, *PC2*, *PC3*, ..., *PC5376*.

La tercera salida es *variances*, un vector que contiene la varianza explicada por el PC correspondiente. Cada columna de *scores* tiene una varianza de la muestra igual a la varianza de los elementos correspondientes. Para observar la representación grafica de la varianza de los PCs, se calcula el porcentaje de la variabilidad total explicada por cada PC del conjunto de datos de *Training* de *"Ia"*.

En la Figura 4.1 se muestra la varianza de los 10 primeros PCs. Como se puede observar, hay un cambio significativo en la cantidad de varianza explicada por el primero y el segundo componente.

El primer componente por sí mismo explica un poco más del 45 % de la varianza, de tal forma que se necesitan más componentes. Los 3 primeros PCs explican cerca de dos tercios de la variabilidad total y los diez 10 PCs explican aproximadamente el 95 % de la variabilidad total de las clasificaciones estandarizadas, por lo que podría ser una manera razonable para reducir las dimensiones con el fin de visualizar los datos.

El resultado de la clasificación en el escenario *offline* y en el escenario *online simulado* para los conjuntos de datos "*Ia*" y "*Ib*" con el vector de

características conformado por los PCs, se muestra en el Capítulo 5.

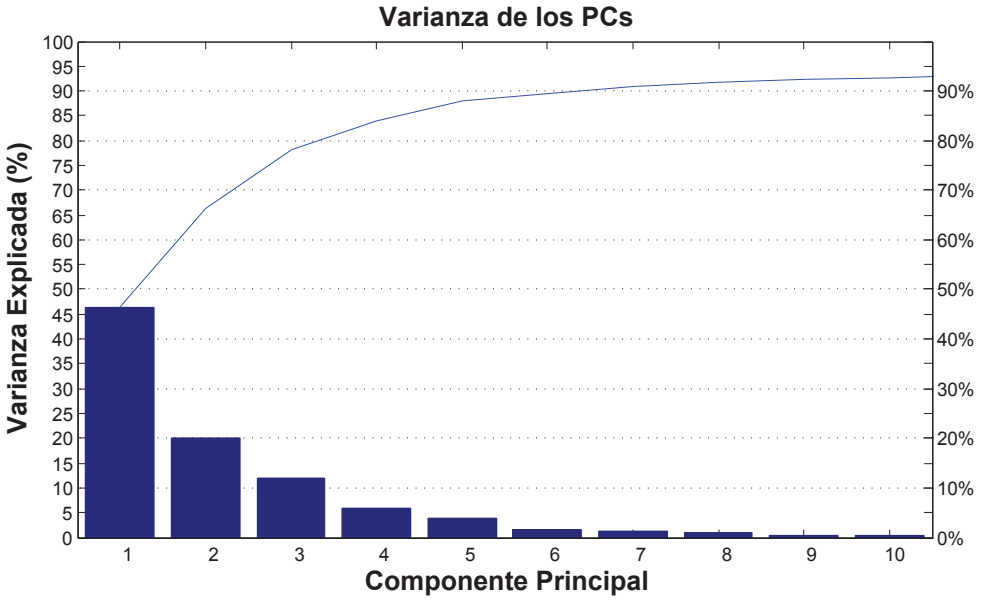

Figura 4.1: *Porcentaje de variabilidad total de cada componente principal.*

4.2.2. Estimación espectral

La mayoría de las señales biomédicas son el resultado de procesos que tienen lugar en el dominio del tiempo. Para el caso de las señales de control SCP, es conveniente realizar su estudio en el dominio de la frecuencia, por lo cual es necesario obtener la función de Densidad Espectral de Potencia (*Power Spectral Density*, PSD).

En la literatura científica, se encuentran distintos métodos para la obtención mediante PSD de una señal. Se pueden distinguir tres grandes clases: (1) Métodos no paramétricos basados en la DFT, (2) Métodos paramétricos basados en modelos ARMA y variantes y (3) Métodos de tiempo-frecuencia que asumen la no estacionariedad de las señales con el tiempo [64].

1. Métodos no paramétricos: Están basados en la Transformada de Fourier discreta (*Discrete Fourier transform*, DFT). Se utiliza de forma intensiva la aproximación directa (*Periodograma*), que se obtiene como la magnitud al cuadrado de la DFT obtenida mediante la Transformada de Fourier Rápida (*Fast Fourier transform*, FFT) y aplicada directamente sobre los datos (previamente enventanados). Los métodos

paramétricos son computacionalmente eficientes y el PSD obtenido es directamente proporcional a la potencia de las componentes espectrales del proceso [64].

En estos modelos, la autocorrelación se estima directamente de la señal. La versión mejorada del método del Periodograma se conoce como el Método de Welch. Una estrategia más moderna es el MTM, siendo éste, el método utilizado por el ganador del conjunto de datos "*Ia*" de SCP [11].

2. Métodos paramétricos: Se conoce que la autocorrelación responde a una determinada forma paramétrica o modelo. Los modelos Autorregresivos (AR) es un método paramétrico, que calcula la PSD estimando primero los coeficientes de un hipotético sistema lineal que genera la señal [96] junto a los Modelos Autorregresivos de Media Móvil (*Auto-Regressive Moving Average Models*, ARMA) y los modelos de Medias Móviles (MA).

   Éstos modelos proporcionan estimadores con mejores características estadísticas, aunque pueden ser computacionalmente menos eficientes que los basados en la DFT [64]. Así mismo, tienden a producir mejores resultados que las técnicas clásicas, cuando la longitud de la señal disponible es relativamente corta [96].

3. Métodos de tiempo-frecuencia: Tanto los métodos paramétricos como los no paramétricos hacen la suposición de que la señal es estacionaria. En caso contrario, se debe implementar un esquema de segmentación que asegure una ventana de datos libre de transitorios [64]. Los métodos de tiempo-frecuencia, asumen la no estacionariedad de las señales con el tiempo. La STFT y la WT son los más usuales en aplicaciones biomédicas.

4.2.2.1. Modelos Autorregresivos

Un modelo Autorregresivo (AR) es un modelo paramétrico lineal que describe la muestra actual como una combinación de las p muestras anteriores. Se expresa como:

$$s(t) = \sum_{i=1}^{p} coef_i s_{t-i} + e(t) \tag{4.3}$$

donde $s(t)$ es la señal de EEG, p es el orden del modelo, $coef_i$ son los parámetros AR que se pretende obtener y $e(t)$ es un término de error que representa

un ruido blanco de media cero y varianza finita.

En AR, el orden del modelo expresa el número de observaciones retrasadas de la serie temporal analizada. La estimación del orden p es muy importante, ya que debe representar correctamente el proceso minimizando a la vez la complejidad del modelo. Para estimar p se puede usar diferentes técnicas de optimización [64]. Para estimar el orden del modelo en las señales de SCP, se toma como referencia [35], [27], [38], [30], variando $p = 1, 2, 3, 4, 5, 6, 8, 10$.

Existen diversos algoritmos para obtener los parámetros del modelo AR. Los más utilizados son: Forward-backward, Least-squares, Geometric lattice, Burg y Yule-Walker [80]. Con un estimador paramétrico basado en el modelo de Yule-Walker, se buscan los parámetros del modelo AR en las señales SCP.

4.2.2.1.1. Modelo de Yule-Walker

El modelo de estimación de Yule-Walker es también llamado Método de Autocorrelación. Consiste en plantear el sistema de ecuaciones de Yule-Walker y proceder a su resolución sustituyendo en dicho sistema las autocorrelaciones por sus estimaciones. Por tanto, se iguala momentos teóricos con estimados.

Si la serie tiene estructura AR(p) que es equivalente a la Ecuación 4.3:

$$s(t) = \phi_1 s_{t-1} + ... + \phi_p s_{t-p} + e(t) \tag{4.4}$$

las ecuaciones de Yule-Walker se obtienen calculando las covarianzas o correlaciones de s_t con s_{t-k} ($k \geq 1$) con lo que obtenemos la ecuación en diferencias:

$$\begin{cases} \gamma_k = \phi_1 \gamma_{k-1} + ... + \phi_p \gamma_{k-p} \\ \qquad\qquad o \\ \rho_k = \phi_1 \rho_{k-1} + ... + \phi_p \rho_{k-p} \end{cases} \tag{4.5}$$

Como estas funciones son pares, podemos plantear un sistema de p ecuaciones con p incógnitas. Al resolverlas obtenemos la estimación de los parámetros ϕ_i, sustituyendo los valores de las covarianzas o correlaciones teóricas por sus estimaciones muestrales.

El valor de la varianza de e se obtiene de la ecuación:

$$\phi_e^2 = \gamma_0 - \hat{\phi}' \begin{pmatrix} \gamma_1 \\ ... \\ \gamma_p \end{pmatrix} \tag{4.6}$$

ecuación para $(k = 0)$.

Las covarianzas del modelo teórico así obtenido coinciden con las muestrales para los valores $k = 0, 1, ..., p$.

En los estudios llevados a cabo en [107], se concluye que el uso de los parámetros AR en señales de EEG es más adecuado que el de los coeficientes de Fourier. También los modelos AR se utilizan frecuentemente debido a su simplicidad y a la existencia de algoritmos efectivos para la estimación de parámetros.

En MatLab, hemos utilizado la función ***aryule*** para estimar los parámetros del modelo AR con el método de Yule-Walker. La sintaxis de la función ***aryule*** es:

[ar_coeffs] = aryule(data,order)

donde ar_coeffs son los parámetros AR que devuelve la función ***aryule***. Dicha función recibe dos parámetros de entrada, *data* (un vector que contiene los datos un canal (896 muestras) del conjunto de *Training* de "Ia") y el parámetro *order* que hace referencia al orden del modelo. El valor de *order* es un número entero positivo que no puede superar la longitud de los datos de entrada.

El resultado de la clasificación en el escenario *offline* y en el escenario *online simulado* para los conjuntos de datos "Ia" y "Ib" con el vector de características conformado por los parámetros AR, se presenta en el siguiente Capítulo.

4.2.2.2. Transformada Wavelet

La Transformada Wavelet (*Wavelet Transform*, WT) representa el paso lógico siguiente a la STFT. Una técnica mediante ventanas con regiones de tamaño variable.

El análisis Wavelet permite el uso de intervalos grandes de tiempo en aquellos segmentos en los que se requiere mayor precisión en baja frecuencia, y regiones más pequeñas donde se requiere información en alta frecuencia. Para ello se emplea una ventana modulada con diferentes escalas en tiempo. Esta idea es la que se muestra en forma esquemática en la Figura 4.2.

Figura 4.2: *Esquema de la Transformada Wavelet. Fuente: [75].*

Una Wavelet es una señal (o forma de onda) de duración limitada cuyo valor medio es cero. Comparando las Wavelets con las funciones sinusoidales (que son la base del análisis de Fourier), se puede resaltar que la principal diferencia radica en que las señales sinusoidales no tienen duración limitada, dado que se extienden desde $-\propto$ a $+\propto$. Además, mientras las señales sinusoidales son suaves y predecibles, las Wavelets tienden a ser irregulares y asimétricas, tal como se puede apreciar en la Figura 4.3 [75].

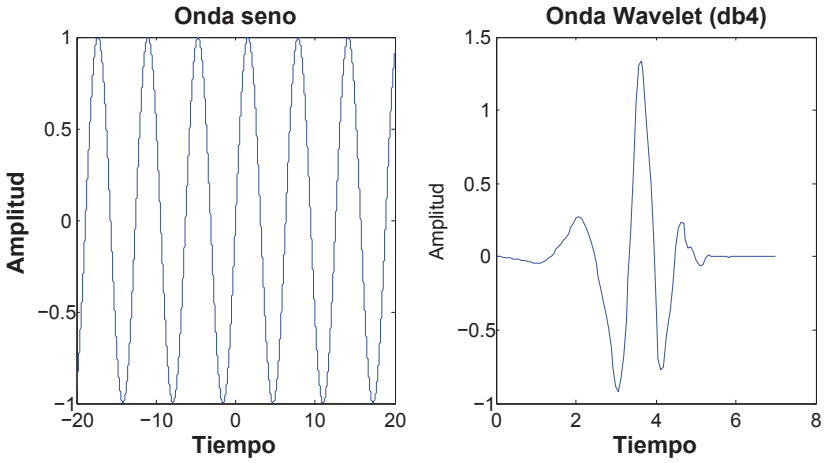

Figura 4.3: *Ejemplo de señal sinusoidal y una señal Wavelet.*

Las Wavelets se emplean para hacer representaciones en tiempo-frecuencia de una señal partiendo de una wavelet madre $\psi(t)$ que se escala por un valor s y se desplaza un tiempo b.

4.2.2.2.1. Transformada Wavelet Continua

La Transformada Wavelet Continua (*Continuous Wavelet Transform*, CWT) de una señal *x(t)* se define como la convolución de la señal con la wavelet conjugada, desplazada y escalada $\psi_{s,b}(t)$:

$$w_c(s,b) = \int_{-\propto}^{\propto} x(t)\psi^*_{s,b}(t)dt \tag{4.7}$$

donde $w_c(s,b)$ son los coeficientes de la transformada wavelet que se corresponde con el escalado por s y desplazamiento en tiempo b. El símbolo '*' expresa la conjugación compleja. La función wavelet $\psi_{s,b}(t)$ es la versión dilatada y desplazada de la función wavelet madre $\psi(t)$:

$$\psi_{s,b}(t) = \frac{1}{\sqrt{s}}\psi\frac{t-b}{s} \tag{4.8}$$

Al permitir variaciones continuas de desplazamientos b y escalados s de la wavelet madre, la CWT implica una carga computacional importante, además de redundancias en los resultados. Esto da lugar a la DWT, que se basa en dividir una señal en sus diferentes componentes frecuenciales y estudiar cada componente con una resolución acorde a su escala.

4.2.2.2.2. Transformada Wavelet Discreta

La Transformada Wavelet Discreta (*Discrete Wavelet Transform*, DWT) es una herramienta de gran utilidad para la caracterización espectral de una señal, pues al utilizar bancos de filtros y hacer descomposiciones en bandas de frecuencia ortogonales, se puede analizar el contenido de energía de una señal en cada una de ellas por separado.

En los datos de SCP, no se emplea bandas de frecuencias específicas, y por ello, se realiza una descomposición de un solo nivel. La señal bruta se filtra mediante una pareja de filtros, uno paso bajo y otro paso alto, que realiza

posteriormente una separación de la señal en dos segmentos. Éste proceso da lugar a unos coeficientes de Aproximaciones (cA) que corresponde a las bajas frecuencias y unos coeficientes de Detalles (cD) que corresponde a las altas frecuencias. El proceso de descomposición wavelet de la señal en un nivel se muestra en la Figura 4.4.

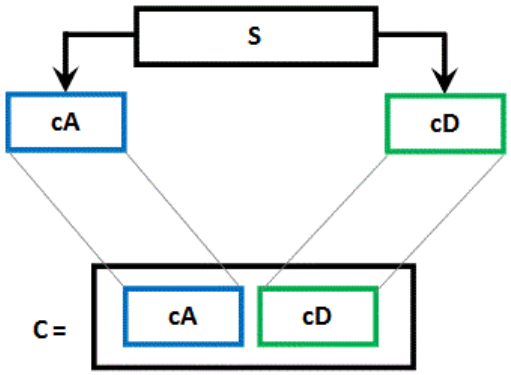

Figura 4.4: *Descomposición Wavelet de la señal generada en un nivel.*

Considerando que en la mayoría de las señales cerebrales, las componentes de baja frecuencia otorgan la mayor parte de su información, o bien, le dan una especie de identidad, el vector de características inicial se forma a partir dichas frecuencias obtenidas.

Para implementar este esquema de descomposición, utilizamos la función ***dwt*** de MatLab con 1 nivel de profundidad y una onda Wavelet Daubechies (db4) (ver Figura 4.3) tomando como referencia el estudio citado en [106]. La sintaxis de la función ***dwt*** es:

[A,D] = dwt(data,'db4')

donde A y D son los coeficientes de Aproximaciones y coeficientes de Detalles que devuelve la función ***dwt***. Dicha función recibe dos parámetros de entrada, *data* (un vector que contiene los datos un canal (896 muestras) del conjunto de *Training* de "Ia") y el parámetro '*db4*', entre comillas simples, que hace referencia al tipo de onda Wavelet que se usa.

El resultado de la clasificación en los escenarios *offline* y *online simulado* para los conjuntos de datos "*Ia*" y "*Ib*" con el vector de características conformado por los coeficientes de Aproximaciones (cA), se muestra en el Capítulo 5.

4.3. Clasificación de características

Una vez caracterizada la señal EEG de estudio, se usa Análisis Discriminante Lineal y una Máquina de Vectores de Soporte, con dos tipos de *kernel*: Lineal y RBF, variando los valores de los parámetros C y γ.

4.3.1. Análisis Discriminante

El Análisis Discriminante (*Discriminant Analysis*, DA) es una de las técnicas de análisis multivariante más conocidas cuyo objetivo es encontrar la combinación lineal (o cuadrática) de las variables independientes (también llamadas variables de clasificación en las que suponemos que se diferencian los grupos) que mejor permitan diferenciar (discriminar) a las grupos (clases), por medio de una función discriminante, la cual podrá ser utilizada para clasificar nuevos casos.

Esta técnica es capaz de aprovechar las relaciones existentes entre una gran cantidad de variables para maximizar la capacidad de discriminación [34].

Existen distintas técnicas de análisis discriminante, tales como Análisis Discriminante Lineal (*Linear Discriminant Analysis*, LDA), Análisis Discriminante Cuadrático (*Quadratic Discriminant Analysis*, QDA) entre otras.

Tanto LDA como QDA parten de una serie de suposiciones. Ambas requieren que las variables que describen las observaciones de cada grupo siga la distribución normal multivariante. Además, la discriminación lineal requiere que la matriz de varianzas-covarianzas de los grupos sean iguales. Este supuesto de homogeneidad de las matrices de varianza-covarianza se cumple raras. Al no cumplirse y al existir normalidad entre los datos, la discriminación cuadrática será más eficiente que la lineal para muestras grandes [70].

LDA y QDA son dos clasificadores clásicos que como su nombre lo indica, tienen una frontera de decisión lineal y cuadrática (elipses, hipérbolas), respectivamente. Estos clasificadores son interesantes porque pueden ser fácilmente calculados, son multiclase y han demostrado funcionar bien en la práctica [13].

La Figura 4.5 muestra las fronteras de decisión para LDA y QDA. La fila de abajo demuestra que LDA solamente puede aprender límites lineales, mientras que QDA puede aprender límites cuadráticos y de ésta forma ser más flexible.

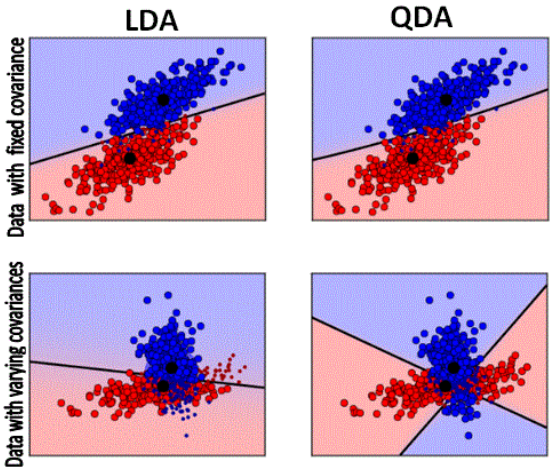

Figura 4.5: *fronteras de decisión para LDA y QDA. Fuente:[13]*

En el diseño de sistemas BCI *online*, LDA proporciona una respuesta rápida con recursos computacionales limitados. LDA es usado para clasificar los patrones encontrados en las señales cerebrales en dos clases, no obstante es posible extender el método para diferenciar entre cualquier número de clases (multiclases).

En la Figura 4.6 están representadas en el espacio definido por las variables X_1 y X_2, las nubes de puntos correspondientes a dos clases que se diferencian entre sí en ambas variables aunque se solapan en una pequeña región situada entre ambas.

Así mismo, está representada la función D, que es una combinación lineal de ambas variables. Sobre la función D se representa la proyección de las dos nubes de puntos en forma de histograma, como si la función D cortara a las dos nubes de puntos en la dirección de su eje. Las dos líneas punteadas de cada uno de los histogramas representan la ubicación proyectada de los puntos medios de cada una de las clases (los centroides).

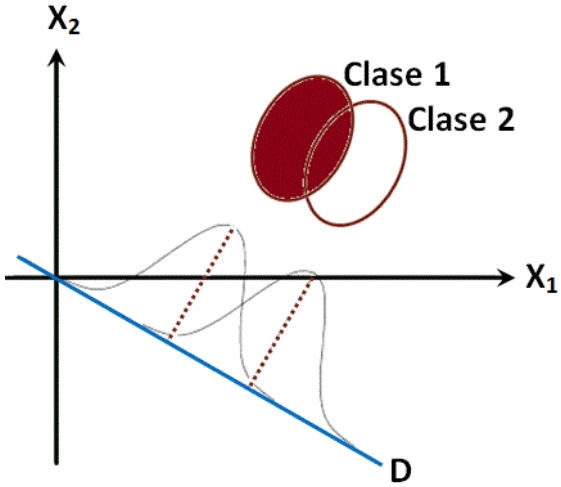

Figura 4.6: *Diagramas de dispersión de dos clases en dos variables de clasificación.*

El propósito de LDA es aprovechar la información contenida en las variables independientes para crear una función D combinación lineal de X_1 y X_2 capaz de diferenciar lo más posible a ambas clases. La función discriminante es de la forma:

$$D = b_1 X_1 + b_2 X_2 \tag{4.9}$$

donde b_1 y b_2 son las ponderaciones de las variables independientes que consiguen hacer que una de las clases obtengan puntuaciones máximas en D, y la otra clase puntuaciones mínimas.

En la Figura 4.7 se representa la función discriminante D extraída del espacio de las variables X_1 y X_2. Las clases aparecen representadas por sus histogramas y las proyecciones de los centroides aparecen marcadas por líneas de puntos.

Sustituyendo en la función discriminante el valor de las medias de la "clase 1" en las variables X_1 y X_2, obtenemos el centroide de la "clase 1":

$$\bar{d}_1 = b_1 \bar{x}_1^{(1)} + b_2 \bar{x}_2^{(1)} \tag{4.10}$$

De igual modo, sustituyendo las medias de la "clase 2", obtenemos el centroide de la "clase 2":

$$\bar{d}_2 = b_1 \bar{x}_1^{(2)} + b_2 \bar{x}_2^{(2)} \tag{4.11}$$

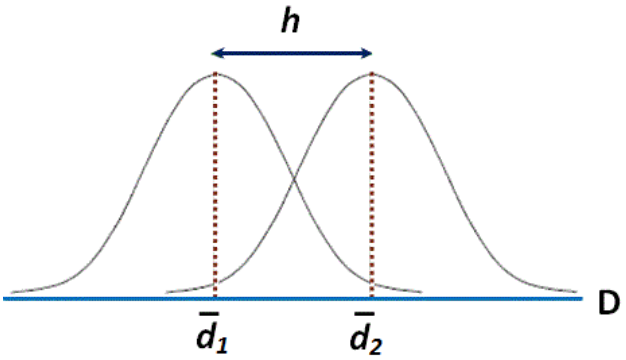

Figura 4.7: *Histogramas de cada clase y centroides representados sobre la función discriminante.*

La función D debe ser tal que la distancia d entre los dos centroides sea máxima, consiguiendo de esta forma que las clases estén lo más distantes posible. Podemos expresar esta distancia de la siguiente manera:

$$h = \bar{d}_1 - \bar{d}_2 \tag{4.12}$$

donde $\bar{d}_1$ y $\bar{d}_2$ son las medias de la "clase 1" y de la "clase 2" en la función D.

Es importante señalar que para un problema de dos clases, LDA supone que las dos clases son linealmente separables como se muestra en la Figura 4.8. Un problema es linealmente separable cuando para cualquier conjunto

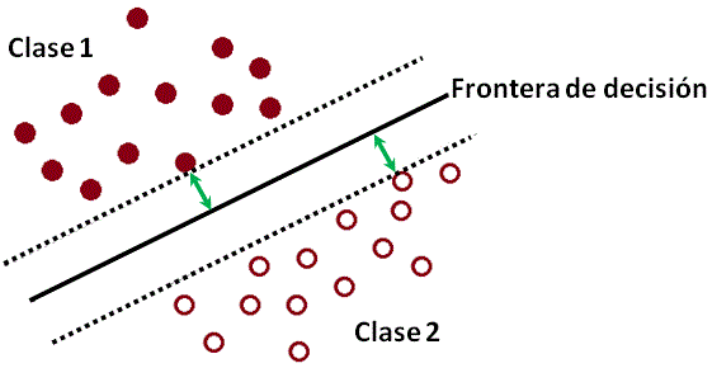

Figura 4.8: *Caso binario linealmente separable. El margen del clasificador lineal es la distancia mínima de cualquier punto a la frontera de decisión.*

de muestras existe un único hiperplano que clasifica con error cero. Al ser el problema linealmente separable, se define una función lineal D como se explicaba anteriormente, que representa un plano en el espacio de características para distinguir la clase a la que pertenece el vector de características a clasificar en función del lado del plano en el que se encuentra el vector.

También es importante resaltar que las clases se deben diferenciar de antemano en las variables independientes ya que el análisis busca diferenciar los grupos al máximo combinando estas variables. Si estas variables no difieren entre los grupos, entonces el análisis será infructuoso, es decir no se podrá encontrar una dimensión en que los grupos se diferencian [34].

Para llevar a cabo la clasificación con LDA, hemos utilizado la función ***classify*** implementada en MatLab, que clasifica cada fila de *sample* en uno de los grupos en *training*. Su sintaxis es de la forma:

class = classify (sample, training, group, 'type')

en donde, la salida *class* indica el grupo al cual cada fila de *sample* ha sido asignada, siendo del mismo tipo que el de *group*.

Los parámetros de entrada *sample* y *training*, son matrices con el mismo número de columnas. *sample* contiene los datos con las muestras con las que se valida el clasificador y *training* los datos de entrenamiento.

El parámetro de entrada *group*, es una variable de agrupación para *training*. Estos valores únicos definen los grupos; cada elemento define el grupo al cual pertenece la correspondiente fila de *training*. *group* es un vector columna. Tanto *group* como *training* tienen el mismo número de filas.

La función *classify* puede realizar la clasificación utilizando diferentes tipos de análisis discriminante. Mediante el parámetro *'type'* dentro de comillas simples, especificamos el tipo de función discriminante que queremos utilizar, tales como, lineal, cuadrática o mahalanobis.

Para clasificar los datos de SCP, utilizamos el Análisis Discriminante Lineal (LDA) como se describió y menciono anteriormente. La función Lineal implementada en Matlab es ***'linear'*** escrita entre comillas simples.

4.3.2. Máquina de Vectores de Soporte

Las Máquinas de Vectores de Soporte (*Support Vector Machine*, SVM) se emplean tanto en aplicaciones de clasificación como en regresión y se basa en la determinación del hiperplano que da lugar a la máxima distancia de separación entre los vectores transformados.

Esta distancia de separación se obtiene mediante la construcción de dos hiperplanos paralelos al hiperplano de separación óptimo, localizados a ambos lados del mismo y que contengan al menos a uno de los vectores transformados, denominados Vectores de Soporte (*Support Vectors*), tal como se muestra en la Figura 4.9; se asume que cuanto mayor sea esta distancia, mejor será la capacidad de generalización del clasificador.

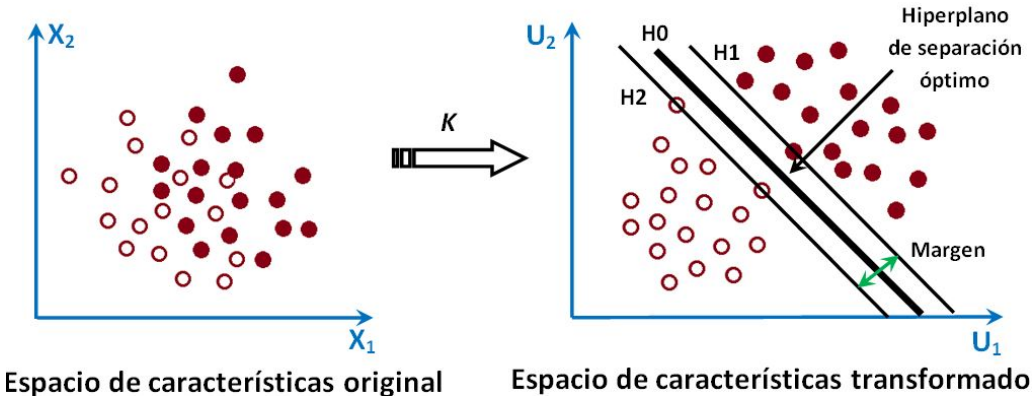

Figura 4.9: *Funcionamiento de una SVM*

Las operaciones que realiza una SVM en clasificación son:

- Transformación de los datos o vectores de características de entrada a un espacio de mayor dimensión a través de una función núcleo ó *kernel* K.
- Cálculo del hiperplano óptimo que maximiza la distancia entre las clases consideradas. Si los datos son linealmente separables, el hiperplano obtenido maximiza el margen de separación, a la vez que minimiza la función de penalización que considera las clasificaciones incorrectas.

En los procesos de clasificación puede ocurrir que los datos no sean linealmente separables o que exista un cierto nivel de ruido. Según esto se pueden

emplear distintos tipos de SVM: (1) SVM lineal con margen máximo, (2) SVM con margen blando o (3) SVM para la clasificación no lineal.

1. SVM lineal con margen máximo.

Sólo se debe emplear cuando los datos son linealmente separables, es decir, cuando es posible trazar una frontera lineal para delimitar ambas clases de datos.

Las SVM conforman hiperplanos que separan los datos de entrada en dos subgrupos que poseen una etiqueta propia. En medio de todos los posibles planos de separación de las dos clases etiquetadas, existe sólo un hiperplano de separación óptimo, H_0, (ver Figura 4.9), que actuando como frontera de separación entre las clases consideradas, maximiza la distancia a los ejemplares más próximos de dichas clases, denominándose a estos datos vectores de soporte, sobre los mismos se dispondrán dos hiperplanos, H_1 y H_2, paralelos al hiperplano H_0. Cuanto mayor sea la distancia mejor será en general el error del clasificador.

Aquellos puntos sobre los cuales se apoya el margen máximo son los denominados vectores de soporte. Un ejemplo de este caso se puede observar en la Figura 4.10 en el que el hiperplano óptimo es aquel que tiene margen 1 porque maximiza el margen. Como se observa, los Vectores de Soporte son aquellos que se encuentran justo a la distancia del margen.

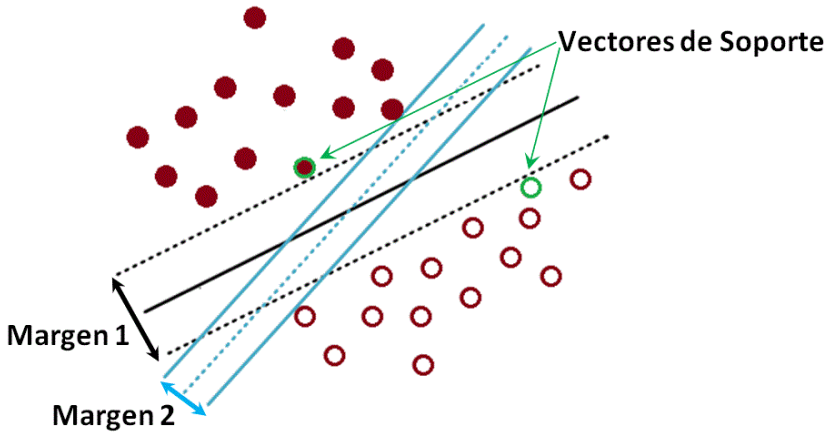

Figura 4.10: *SVM linealmente separable. Selección del hiperplano de separación óptimo*

2. SVM con margen blando

En el análisis anterior se considera que en la región comprendida entre los hiperplanos H_1 y H_2 no existen vectores, ni que éstos pudieran ser erróneamente clasificados.

Esta consideración es muy exigente para su aplicación en casos prácticos, lo que motivó una nueva idea de margen máximo, denominada "Margen Blando" la cual, para el caso en que no sea posible encontrar un hiperplano de separación óptimo, H_0, seleccionara aquel hiperplano que separando los vectores de entrada tan limpiamente como fuera posible, permitiera errores de clasificación a la vez que maximiza la distancia al resto de los datos que han sido correctamente identificados, tal como se muestra en la Figura 4.11.

Esta nueva variante introduce un nuevo conjunto de variables, denominadas "variables de holgura": $h_i, i = 1, ..., N$, que permiten ponderar el grado de desclasificación de cada vector.

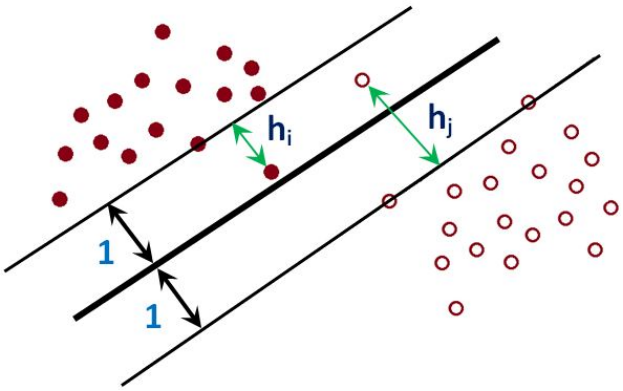

Figura 4.11: *SVM linealmente separable. Margen Blando.*

3. SVM para la clasificación no lineal

Para un problema no separable linealmente, puede aplicarse una transformación a través de una función no lineal K de los conjuntos de datos originales a conjuntos de datos en espacios de características transformados, de dimensión superior a la del espacio de características original, y separables linealmente, emplearan el algoritmo de determinación de hiperplano óptimo, H_0, como método de separación de dichos conjuntos.

De este modo, aunque el clasificador se base en un hiperplano en el espacio de características transformado, este puede ser no lineal en el espacio de características original.

Las funciones que se usan para poder realizar la transformación de los datos de un espacio inicial a otro de mayor dimensión se logra mediante el uso de una función núcleo ó *kernel*. De manera gráfica se puede observar en la Figura 4.12 como la función *kernel* permite realizar la separación y el traslado de los datos al espacio de características.

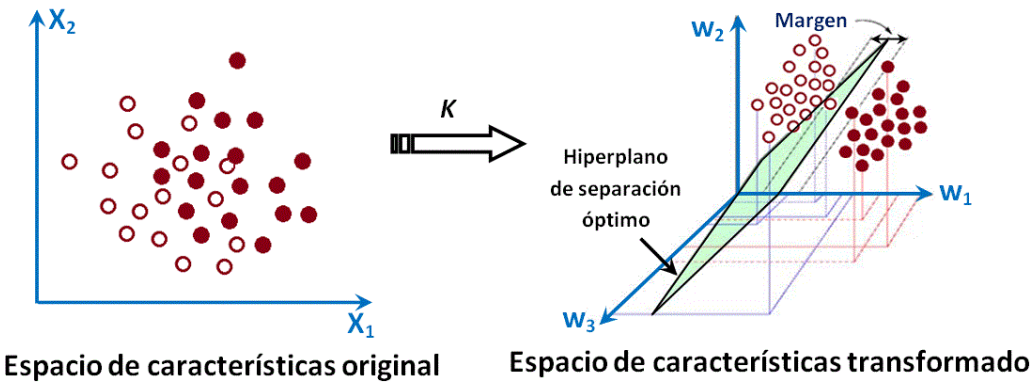

Figura 4.12: *SVM no linealmente separable. Transformación de los datos de entrada a un espacio de mayor dimensión inducida por una función kernel.*

Una Función *kernel* $k(x, y)$ es un producto interno en el espacio de características, que tiene su equivalente en el espacio de entrada [49].

Entre los *kernels* más comunes, se encuentran: la función lineal, la función polinomial, la función de base radial (*Radial Basis Function*, RBF), entre otras.

- Lineal:

$$k(x, y) = x * y \tag{4.13}$$

- Polinómica:

$$k(x, y) = (\gamma x * y + c)^{\alpha} \tag{4.14}$$

  siendo c un coeficiente y α el rango del polinomio.

- RBF:

$$k(x, y) = exp(-\gamma \mid x - y \mid^2) \tag{4.15}$$

  siendo γ una constante de proporcionalidad.

Según el tipo de *kernel* y de los valores de sus parámetros, se puede obtener distintas fronteras de decisión. El *kernel* que se utiliza generalmente en el campo de los sistemas BCI es RBF, que ha sido aplicado para clasificar a los Potenciales Evocados P300 [84].

Para llevar a cabo la clasificación de los datos de SCP, empleamos la función Kernel RBF que además presenta menos dificultades numéricas, ya que los valores de salida de las funciones gaussianas se encuentran entre 0 e $\propto$ [108].

Para resolver una tarea de clasificación con SVM y una función *kernel* RBF, hay que determinar dos parámetros principales:

1. C: Parámetro de regularización que controla el equilibrio entre la maximización del margen y la clasificación del conjunto de entrenamiento sin error.

2. γ: Parámetro de suavizado que determina el ancho de la gaussiana, y con ello la sensibilidad de la medición de la distancia.

Estos parámetros se calculan a través de la selección de distintos valores para C y γ evaluando de forma empírica el resultado con todas las combinaciones posibles. En [56], los autores recomiendan una "red de búsqueda" para C y γ utilizando validación cruzada. Por medio de un método bastante práctico, descubrieron que tomando valores para C y γ en secuencias de crecimiento exponencial se pueden identificar buenos pares de parámetros.

Para crear y entrenar la SVM hemos utilizado la función ***svmtrain*** de Matlab que tiene la siguiente estructura:

svmstruct = svmtrain (training, train_class, 'kernel_function', 'boxconstraint')

donde,

- *training*: Es un matriz con los datos de entrenamiento.

- *train_class*: Especifica las etiquetas (0/1) de entrenamiento.

- *kernel_function*: Es la función *kernel* que se va a usar. Los tipos de función *kernel* implementados son: lineal, cuadrática, polinómica, gaussiana, entre otras.

  Para la clasificación de los datos de SCP con una SVM, se usará dos tipos de *kernel* diferentes:

  - ◇ Función Lineal: En Matlab, la función es ***'linear'*** escrita entre comillas simples.
  - ◇ Función Gaussiana de Base Radial: La función es ***'rbf'*** en comillas simples. Para entrenar la SVM, usamos diferentes valores para γ, siguiendo una secuencia de crecimiento exponencial. Los valores de γ los definimos en el rango: $\gamma = 2^{-15}$, $\gamma = 2^{-13}$,...,$\gamma = 2^{7}$.

- 'boxconstraint' o parámetro de suavizado C: El parámetro C va a tomar distintos valores entre sí, para ver como varía el resultado, siguiendo el mismo método utilizado para obtener los valores de γ. Los valores que toma C son: $C = 2^{-7}$, $C = 2^{-5}$,...,$C = 2^{15}$.

Para probar el desempeño del clasificador utilizamos ***svmclassify*** que recibe como parámetros la SVM creada anteriormente y los datos de prueba (Testing).

Capítulo 5

Resultados y discusión

En éste capítulo se incluye la discusión de los resultados de la clasificación obtenidos con las diferentes técnicas estadísticas propuestas en el Capítulo 3, así como los resultados de la clasificación después de aplicar los diferentes métodos de extracción de características explicados en el Capítulo anterior de los conjuntos de datos "*Ia*" y "*Ib*".

Los resultados de la clasificación en el escenario *offline* se presenta en Tablas, en las que se describe el porcentaje de error obtenido tanto para la "clase 0" como para la "clase 1" en ambos conjuntos de datos utilizando LDA y una SVM. Por medio de representaciones gráficas, se muestra los resultados de la clasificación en el escenario *online simulado*.

5.1. Métodos de evaluación de desempeño

La evaluación de un algoritmo de construcción de modelos de clasificación se puede realizar atendiendo a distintos aspectos, tales como, precisión, eficiencia, robustez, escalabilidad, interpretabilidad, complejidad entre otros. Existen diferentes métricas y métodos que permiten evaluar y estimar de forma fiable, la "calidad" de un modelo de clasificación.

Es necesario establecer una métrica común para evaluar el desempeño o la precisión del clasificador en los sistemas BCI. Existen diferentes métricas, tales como, la Función o Matriz de Costes, la Matriz de Confusión, las Curvas ROC y DET, la Velocidad de Transferencia de Información (*Information Transfer Rate*, ITR), entre otras.

La forma práctica de evaluar los clasificadores en este TFM, es elaborando una matriz de confusión. La matriz de confusión es una herramienta del análisis estadístico multivariante, el cual propone crear una matriz de $n * n$, donde n es el número de clases, como se muestra en la Figura 5.1.

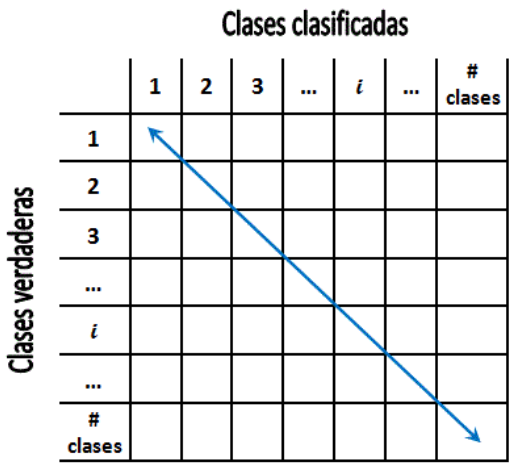

Figura 5.1: *Estructura de una matriz de confusión*

La forma de llenar esta matriz, primero se tiene que tener un conjunto de entrenamiento (*Training*) y un conjunto de datos de validación o prueba (*Testing*), de las cuales conocemos a priori su clase. Una vez que el clasificador ha sido entrenado con el conjunto de entrenamiento, se le envían las muestras del conjunto de prueba, para así obtener un resultado de clasificación de dichas muestras.

Si la muestra n con clase i es clasificada en la i-esima clase entonces se coloca en las coordenadas i, j y el resultado se considera correcto, pero en caso de que la muestra sea clasificada en la clase j, entonces se considera un error y se coloca dentro de la matriz en la posición i, j, por lo que el número de aciertos estará posicionado sobre la diagonal de la matriz y los errores estarán dispersos por el resto de ella.

De esta forma, la matriz de confusión no solo permite observar y contar fácilmente los errores, si no permite saber exactamente en qué situaciones el clasificador está fallando y con qué clase se está confundiendo.

5.2. Resultados obtenidos

Los resultados obtenidos, han sido alcanzados utilizando dos metodologías diferentes de clasificación:

1. Se ha comenzado por simular con el conjunto de entrenamiento (*Training*), realizando tres diferentes tipos de particiones de datos (P1, P2 y P3 tal y como se muestra en la Tabla 5.1) para entrenar y validar el modelo.

Tipos de particiones de *Training*		
Partición	Entrenar	Validar
P1	25 %	75 %
P2	50 %	50 %
P3	75 %	25 %

Tabla 5.1: *Particiones de los datos de SCP del conjunto de Training.*

   Con ello, pretendemos hacer un análisis más exhaustivo de los datos de SCP, con el fin de predecir y estimar cómo de preciso es nuestro modelo para llevar a cabo una posterior clasificación lo mejor posible.

   Utilizaremos la técnica de la validación cruzada aleatoria para evaluar los resultados obtenidos con el conjunto de entrenamiento (*Training*) y garantizar que son independientes de la partición entre datos de entrenamiento y validación. Esta técnica consiste principalmente en realizar 100 iteraciones de forma aleatoria con cada una de las particiones establecidas y calcular el valor medio de los porcentajes de error obtenidos en cada una de las clases.

   A partir de ahora, le llamamos *Training*, al error de validación obtenido con 100 iteraciones, del conjunto de datos de *Training* que suministra "BCI Competition II".

2. Después se ha procedido a probar los clasificadores con la misma configuración, pero entrenando con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*) iterando una sola vez.

Este proceso se ha seguido tanto para el conjuntos de datos "*Ia*" (sujeto sano) como para el conjunto de datos "*Ib*" (paciente con ALS).

5.2.1. Resultados del conjunto de datos "*Ia*"

A continuación presentamos los resultados de la clasificación del conjunto de datos "*Ia*", obtenidos con el conjunto de entrenamiento (*Training*) y cada una de las particiones establecidas para entrenar y validar el modelo, en los escenarios *offline* y *online simulado*. Seguido, la clasificación con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*), en ambos escenarios de análisis ya mencionados.

5.2.1.1. Resultado de la clasificación en el escenario *offline*, con el conjunto de datos de *Training* particionado.

Se muestra a continuación, el porcentaje de error en la clasificación *offline* utilizando LDA y una SVM con un *kernel* Lineal y las características obtenidas usando las diferentes técnicas de visualización y caracterización propuestas en el Capítulo 3, así como los diferentes métodos de extracción de características estudiados. El vector está conformado por las características extraídas de un método en concreto.

En cada una de las siguientes Tablas, se puede observar el porcentaje de error obtenido en la clasificación tanto para la "clase 0" como para la "clase 1" y el promedio o media de ambas clases. Las particiones P1, P2 y P3 corresponden a las particiones hechas en el conjunto de *Training* para entrenar y validar el modelo.

Las celdas sombreadas de azul, representan los mejores porcentajes de error alcanzados en las clasificaciones.

5.2.1.1.1. Clasificación con el nivel DC

Después de realizar varias pruebas, utilizando los valores del nivel DC obtenidos de diferentes combinaciones de canales, se presenta el menor y mejor porcentaje de error alcanzado. Dicho resultado se obtiene con los canales 1 y 2, tal como propusimos en el Capítulo 3, ya que son los canales que presentan una clara diferencia de separación entre las clases. Los valores del nivel DC del canal 5, no mejora el desempeño del clasificador, por esta razón no se tiene en cuenta.

En la Tabla 5.2, se muestra el porcentaje de error alcanzado en la clasificación con LDA.

Los segmentos DC se refieren a las divisiones de la señal de cada uno de los canales 1 y 2 (cada canal está compuesto por 896 muestras). En cada trozo o segmento de la señal dividida se calcula el nivel DC respectivo.

Como puede verse en la Tabla 5.2, el error en la clasificación con el valor del nivel DC es inversamente proporcional al número de segmentos, de tal forma, que entre más segmentos DC se empleen, el error de la clasificación disminuye progresivamente hasta un máximo de 8 divisiones.

El mejor desempeño de LDA se logra con la partición de datos P3, en el que se entrena con el 75 % de los datos y se valida con el 25 % restante.

Nivel DC - LDA				
Segmentos DC	Partición	Canales: 1 y 2		
		clase 0	clase 1	media cl0 y cl1
1	P1	31,28 %	30,36 %	**30,82 %**
	P2	31,17 %	30,07 %	**30,62 %**
	P3	31,17 %	30,51 %	**30,84 %**
2	P1	24,15 %	28,86 %	**26,50 %**
	P2	24,28 %	27,93 %	**26,10 %**
	P3	24,31 %	27,49 %	**25,90 %**
3	P1	17,58 %	25,61 %	**21,59 %**
	P2	17,82 %	24,56 %	**21,19 %**
	P3	17,31 %	24,06 %	**20,69 %**
4	P1	16,58 %	24,38 %	**20,48 %**
	P2	16,93 %	22,65 %	**19,79 %**
	P3	17,30 %	21,61 %	**19,46 %**
5	P1	15,85 %	23,41 %	**19,63 %**
	P2	15,82 %	22,40 %	**19,11 %**
	P3	15,63 %	22,45 %	**19,04 %**
8	P1	16,42 %	24,61 %	**20,52 %**
	P2	15,24 %	22,74 %	**18,99 %**
	P3	14,62 %	21,66 %	**18,14 %**

Tabla 5.2: *Resultados de la clasificación offline con LDA. Vector de características compuesto por el nivel DC de los canales 1 y 2 para el conjunto de datos de Training con particiones de "Ia".*

La dimensión del vector de características que se obtiene desde los valores del nivel DC, es definido de acuerdo a la cantidad de segmentos en los que se

divide la señal. Por ejemplo, si se obtiene el valor DC de la señal completa para los canales 1 y 2, el vector de características será una matriz de $268 * 2$ y si se obtiene el valor DC de la señal con 8 segmentos, el vector estará conformado por una matriz de $268 * 16$.

Con lo anterior se puede decir, que aunque se incrementen los segmentos para alcanzar un mejor desempeño del clasificador, esto no afecta ni incrementa el tiempo de proceso, ya que obtener los valores del nivel DC de las señales no es costoso computacionalmente.

En la Tabla 5.3 se observa el porcentaje de error en la clasificación con LDA, utilizando un vector de características conformado por los valores del nivel DC y adicionando la medida de la Varianza obtenida de los canales 1 y 2. Una vez más, dicha combinación de canales resulta ser la más representativa.

Sólo se muestra los resultados obtenidos con la partición de datos P3, ya que representa el menor porcentaje error en la clasificación tanto para la "clase 0" como para la "clase 1".

Nivel DC y Varianza - LDA				
Segmentos DC y Varianza	Partición	Canales: 1 y 2		
		clase 0	clase 1	media cl0 y cl1
1	P3	17,80 %	29,04 %	**23,42 %**
2	P3	15,71 %	29,80 %	**22,76 %**
3	P3	15,54 %	25,09 %	**20,31 %**
4	P3	15,45 %	25,23 %	**20,34 %**
5	P3	15,38 %	22,82 %	**19,10 %**
8	P3	17,28 %	23,63 %	**20,46 %**

Tabla 5.3: *Resultados de la clasificación offline con LDA. Vector de características compuesto por el nivel DC y la Varianza de los canales 1 y 2 para el conjunto de datos de Training con particiones de "Ia".*

En la Tabla anterior 5.3, la medida de la Varianza aporta información cuando se toma la señal completa, es decir, cuando ésta no se divide en segmentos. Al realizar la clasificación con el valor del nivel DC, se obtiene un porcentaje de error de 30,84 % que se reduce en un 7,42 % al adicionar el valor de la medida de la Varianza (siendo ésta de 23,42 %).

Al dividir la señal en 2, 3 y 5 segmentos y agregar la medida de la Varianza, se obtiene una leve disminución del porcentaje de error en la clasificación, de tal forma que al realizar la clasificación sólo con el valor del nivel DC, el porcentaje de error es 25,90 %, 20,69 % y 19,11 %, y con la medida de la Varianza, éste disminuye en un 7,42 %, 0,38 % y 0,01 % respectivamente.

En algunos casos, como por ejemplo, al dividir la señal en 4 y 8 segmentos, la medida de la Varianza parece ser un valor redundante o poco significativo.

La dimensión del vector de características obtenido desde los valores DC y la medida de la Varianza, es proporcional a la cantidad de segmentos en los que se divide la señal. El cálculo adicional de la medida de la Varianza incrementa el doble el vector de características.

5.2.1.1.2. Clasificación con la Integral

El vector de características está conformado por los valores de aproximación al valor del área bajo la curva de las señales de SCP. El porcentaje de error en la clasificación para la "clase 0" y para la "clase 1" se presenta en la Tabla 5.4 usando la combinación de los canales 1 y 2, siguiendo nuestro análisis propuesto en el Capítulo 3.

Tal y como se puede observar, los valores de la integral obtenidos de dichos canales no es suficiente o no aporta información significativa para alcanzar un buen desempeño con el clasificador LDA.

Integral - LDA			
Particiones	Canales: 1 y 2		
	clase 0	clase 1	media cl0 y cl1
P1	31,15 %	30,49 %	**30,82 %**
P2	31,27 %	30,33 %	**30,80 %**
P3	31,12 %	30,07 %	**30,60 %**

Tabla 5.4: *Resultados de la clasificación offline con LDA. Vector de características compuesto por el valor de la Integral de los canales 1 y 2 para el conjunto de datos de Training con particiones de "Ia".*

En la Tabla 5.5, se presenta los resultados de la clasificación utilizando la combinación de canales 1, 2 y 5.

Como puede verse, al utilizar los valores de aproximación al valor del área bajo la curva del canal 5, el desempeño del clasificador mejora un 4,48 %, lo que supone, que dicho canal, aporta información significativa.

Integral - LDA			
Particiones	Canales: 1,2 y 5		
	clase 0	clase 1	media cl0 y cl1
P1	25,73 %	28,51 %	**27,12 %**
P2	25,44 %	28,09 %	**26,77 %**
P3	25,02 %	27,22 %	**26,12 %**

Tabla 5.5: *Resultados de la clasificación offline con LDA. Vector de características compuesto por el valor de la Integral de los canales 1, 2 y 5 para el conjunto de datos de Training con particiones de "Ia".*

Finalmente, en la Tabla 5.6, se muestra el porcentaje de error en la clasificación usando todos los 6 canales.

Tal y como se puede observar, para obtener un buen desempeño con los valores de aproximación al valor del área bajo la curva, es necesario la combinación de todos los 6 canales, ya que todos en su conjunto, revelan información significativa para llevar a cabo la clasificación.

Integral - LDA			
Particiones	Canales: 1,2,3,4,5 y 6		
	clase 0	clase 1	media cl0 y cl1
P1	24,86 %	24,10 %	**24,48 %**
P2	24,62 %	22,68 %	**23,65 %**
P3	24,78 %	21,44 %	**23,11 %**

Tabla 5.6: *Resultados de la clasificación offline con LDA. Vector de características compuesto por el valor de la Integral de todos los 6 canales para el conjunto de datos de Training con particiones de "Ia".*

Los menores porcentajes de error alcanzados en la clasificación, usando las 3 diferentes combinaciones de canales: (1) Canales 1 y 2, (2) Canales 1, 2 y 5 y (3) Todos los 6 canales, han sido obtenidos con la partición de datos P3, con una mínima diferencia.

5.2.1.1.3. Clasificación con PCA

Con PCA se logra reducir la dimensionalidad de los datos de SCP y a la vez hallar las causas de la variabilidad de las señales y ordenarlas por importancia según su varianza, de tal manera que *PC1* representa la dirección de la mayor varianza posible, la dirección de *PC2*, representa las más alta varianza restante de *PC1* y así sucesivamente.

Las Tablas que se muestran a continuación, muestran los mejores resultados obtenidos en las clasificaciones con PCA, utilizando la partición de datos P3.

Después de realizar diversas pruebas con los PCs extraídos desde diferentes combinaciones de canales, se presentan aquellas combinaciones con las que se obtienen mejores desempeños en la clasificación, tomando también como referencia los resultados del análisis visual realizado en el Capítulo 3.

La Tabla 5.7 muestra el porcentaje de error en la clasificación utilizando los PCs extraídos de los canales 1 y 2.

PCA - LDA			
PCs	Canales: 1 y 2		
	clase 0	clase 1	media cl0 y cl1
1	30,76 %	28,15 %	**29,45 %**
2	30,45 %	30,33 %	**30,39 %**
3	22,05 %	27,81 %	**24,93 %**
4	11,37 %	21,91 %	**16,64 %**
5	12,03 %	20,59 %	**16,31 %**
6	14,80 %	21,21 %	**18,00 %**
7	14,71 %	21,60 %	**18,15 %**
8	15,23 %	22,22 %	**18,73 %**
9	15,39 %	22,31 %	**18,85 %**
10	16,16 %	21,29 %	**18,73 %**

Tabla 5.7: *Resultados de la clasificación offline con LDA. Vector de características compuesto por los PCs de los canales 1 y 2 para el conjunto de datos de Training con particiones de "Ia".*

Como se puede observar, con los primeros 5 PCs, es posible obtener el menor y mejor porcentaje de error en la clasificación de la "clase 0" y la "clase

1" usando solamente los canales 1 y 2. Con esto se logra reducir considerablemente el porcentaje de error en casi un 50 % a diferencia de los métodos anteriormente descritos.

El vector de características compuesto por los 5 PCs de los dos canales tiene una dimensión de 268*5 a diferencia de la dimensión del vector de características conformado por los valores del nivel DC y la medida de la Varianza sin segmentaciones y los valores de aproximación al valor del área bajo la curva (que tienen una dimensión de 268*2).

En la Tabla 5.8, se puede observar el porcentaje de error en la clasificación utilizando los 6 primeros PCs de los canales 1, 2 y 5.

Como puede verse, el canal 5 que se adiciona, mejora un 0,65 % el desempeño del clasificador LDA. El vector de características está conformado finalmente por un matriz con una dimensión de $268 * 6$.

PCA - LDA			
PCs	Canales: 1,2 y 5		
	clase 0	clase 1	media cl0 y cl1
1	30,20 %	29,40 %	**30,80 %**
2	30,21 %	25,97 %	**28,09 %**
3	24,29 %	26,82 %	**25,55 %**
4	19,63 %	21,41 %	**20,52 %**
5	11,72 %	21,96 %	**16,84 %**
6	13,16 %	18,17 %	**15,66 %**
7	13,91 %	18,08 %	**15,99 %**
8	13,57 %	18,23 %	**15,90 %**
9	14,23 %	17,81 %	**16,02 %**
10	14,21 %	17,28 %	**15,74 %**

Tabla 5.8: *Resultados de la clasificación offline con LDA. Vector de características compuesto por los PCs de los canales 1, 2 y 5 para el conjunto de datos de Training con particiones de "Ia".*

La Tabla 5.9 presenta el porcentaje de error en la clasificación usando LDA y el vector de características compuesto por los PCs de todos los 6 canales.

Como se puede ver, al utilizar un PC más (los primeros 11 PCs) se logra

el menor porcentaje de error y el mejor desempeño del clasificador LDA con los PCA. El vector de características compuesto por los 11 PCs, tiene una dimensión de $268 * 11$.

PCA - LDA			
PCs	Canales: 1,2,3,4,5 y 6		
	clase 0	clase 1	media cl0 y cl1
1	33,47 %	30,98 %	**32,23 %**
2	31,73 %	32,87 %	**32,30 %**
3	30,33 %	27,38 %	**28,85 %**
4	26,66 %	30,07 %	**28,37 %**
5	25,26 %	28,13 %	**26,70 %**
6	17,31 %	16,19 %	**16,75 %**
7	17,50 %	15,40 %	**16,45 %**
8	12,86 %	15,12 %	**13,99 %**
9	12,65 %	14,75 %	**13,70 %**
10	12,89 %	14,72 %	**13,81 %**
11	11,91 %	14,27 %	**13,09 %**

Tabla 5.9: *Resultados de la clasificación offline con LDA. Vector de características compuesto por los PCs de todos los 6 canales para el conjunto de datos de Training con particiones de "Ia".*

De acuerdo a los resultados presentados anteriormente, podemos deducir que la dimensión de los datos es importante en PCA, es decir, entre más información haya (canales), mejor será el proceso de reducción de dicha dimensión.

5.2.1.1.4. Clasificación con AR

En el análisis espectral utilizado en las señales de SCP, se utiliza un estimador paramétrico basado en el modelo de Yule-Walker. Para llevar a cabo la clasificación con dicho modelo, se ha utilizado una SVM con un *kernel* Lineal.

Las suposiciones de normalidad e igualdad de varianza no siempre se cumplen en los datos y cuando esto ocurre, LDA no puede ser usado para llevar a cabo la clasificación, es por ello, que se ha empleado una SVM.

Las Tablas que se muestran a continuación, muestran los mejores resultados obtenidos en las clasificaciones con los parámetros AR, variando el valor

del orden p del modelo, como se estableció en el Capítulo anterior y utilizando la partición de datos P3.

En la Tabla 5.10 se muestra el porcentaje de error en la clasificación usando los parámetros AR de los canales 1 y 2. Estableciendo un orden 5, es posible obtener el menor porcentaje de error en dicha clasificación.

AR - SVM Lineal			
Orden	Canales: 1 y 2		
	clase 0	clase 1	media cl0 y cl1
1	20,43 %	37,04 %	**28,74 %**
2	36,37 %	37,74 %	**37,06 %**
3	27,55 %	29,48 %	**28,52 %**
4	27,92 %	25,50 %	**26,71 %**
5	28,03 %	25,18 %	**26,60 %**
6	30,23 %	25,42 %	**27,82 %**
8	29,68 %	27,32 %	**28,50 %**
10	31,61 %	27,25 %	**29,43 %**

Tabla 5.10: *Resultados de la clasificación offline con una SVM Lineal. Vector de características compuesto por los parámetros AR de los canales 1 y 2 para el conjunto de datos de Training con particiones de "Ia".*

Al utilizar la combinación de canales 1, 2 y 5 y definiendo un orden 5, como se muestra en la Tabla 5.11, es posible obtener un menor porcentaje de error (22,74 %).

Finalmente, en la Tabla 5.12 se muestra los resultados de los porcentajes de error obtenidos en la clasificación, utilizando los componentes AR de todos los 6 canales.

Después de observar los resultados con cada una de las tres combinaciones de canales, se puede decir, que la clasificación llevada a cabo con los parámetros AR obtenidos de los canales 1, 2 y 5 disminuye el porcentaje error de clasificación un 1,11 % y un 3,86 % con respecto a utilizar todos los 6 canales y los canales 1 y 2 respectivamente, lo que indica, que la combinación de los canales 1, 2 y 5 es la que aporta información más relevante para dicha clasificación.

Los vectores de características obtenidos para cada una de las tres combinaciones de los canales, tiene una dimensión de 268∗10, 268∗15 y 268∗30 para los canales 1 y 2, canales 1, 2 y 5 y para todos los 6 canales respectivamente.

AR - SVM Lineal			
Orden	Canales: 1, 2 y 5		
	clase 0	clase 1	media cl0 y cl1
1	30,55 %	32,91 %	**31,37 %**
2	34,43 %	36,31 %	**35,37 %**
3	28,59 %	26,37 %	**27,48 %**
4	23,66 %	22,15 %	**22,90 %**
5	23,26 %	22,23 %	**22,74 %**
6	23,90 %	24,06 %	**23,98 %**
8	29,37 %	27,47 %	**26,92 %**
10	26,42 %	25,75 %	**26,09 %**

Tabla 5.11: *Resultados de la clasificación offline con una SVM Lineal. Vector de características compuesto por los parámetros AR de los canales 1, 2 y 5 para el conjunto de datos de Training con particiones de "Ia".*

AR - SVM Lineal			
Orden	Canales: 1,2,3,4,5 y 6		
	clase 0	clase 1	media cl0 y cl1
1	31,84 %	33,62 %	**32,73 %**
2	36,47 %	37,88 %	**37,18 %**
3	30,07 %	26,03 %	**28,05 %**
4	24,54 %	23,15 %	**23,85 %**
5	25,22 %	23,87 %	**24,54 %**
6	27,46 %	24,17 %	**25,82 %**
8	28,77 %	28,62 %	**27,70 %**
10	28,49 %	27,76 %	**28,12 %**

Tabla 5.12: *Resultados de la clasificación offline con una SVM Lineal. Vector de características compuesto por los parámetros AR de todos los 6 canales para el conjunto de datos de Training con particiones de "Ia".*

5.2.1.1.5. Clasificación con DWT

El vector de características está formado por las componentes de baja frecuencia (cA) extraídas de las señales de SCP mediante la descomposición Wavelet de un solo nivel, utilizando una onda Wavelet Daubechuies (db4). Inicialmente se hacen pruebas con LDA, pero no es posible realizar dicha clasificación, ya que el número de observaciones supera el número de las variables, debido a ello, hemos utilizado una SVM con un *kernel* Lineal.

La Tabla 5.14, muestra los porcentajes de error adquiridos con las tres diferentes combinaciones de canales y la partición de datos P3, con los que se obtuvieron mejores resultados en el desempeño con la SVM.

Como puede verse, hay una disminución del porcentaje de error alcanzado con los cA, proporcional al número de canales utilizados (o combinaciones de canales propuestos en el análisis visual de los datos en el Capítulo 3). El vector compuesto por las componentes de baja frecuencia extraídas de cada uno de los 6 canales, es más significativo, ya que se obtiene el menor porcentaje de error en la clasificación (15,48 %).

DWT - SVM Lineal			
Canales	clase 0	clase 1	media cl0 y cl1
1 y 2	25,80 %	30,21 %	**28,00 %**
1,2 y 5	18,91 %	24,32 %	**21,62 %**
1,2,3,4,5 y 6	13,59 %	17,38 %	**15,48 %**

Tabla 5.13: *Resultados de la clasificación offline con una SVM Lineal. Vector de características compuesto por los coeficientes de Aproximación obtenidos desde la DWT para el conjunto de datos de Training con particiones de "Ia".*

5.2.1.1.6. Resumen de los mejores resultados de la clasificación *offline* con el conjunto de datos de *Training* con P3 de "*Ia*" utilizando características extraídas con un método concreto.

Finalmente, en la Tabla 5.14, se muestra un resumen con los mejores resultados alcanzados en la clasificación *offline* con el conjunto de entrenamiento (*Training*), realizando una partición (P3) del 75 % para entrenar y un 25 % para validar el modelo del conjunto "*Ia*". Dichos resultados, pueden ser utilizados para una posible formación de súper vectores.

Para llevar a cabo la clasificación, se utilizó LDA y una SVM con un *kernel* Lineal, y vectores compuestos por características extraídas con un método en concreto.

Resumen clasificación *offline* con *Training* (P3) de *"Ia"*		
Clasificador	Vector de características	% error
LDA	DC (8 Segmentos)(Canal 1 y 2)	18,14
LDA	DC y Varianza (5 Segmentos)(Canal 1 y 2)	19,10
LDA	Integral (Todos los 6 canales)	23,11
LDA	PCA (11 PCs) (Todos los 6 canales)	13,09
SVM Lineal	AR (Orden 4) (Canal 1, 2 y 5)	22,74
SVM Lineal	DWT (db4) (Todos los 6 canales)	15,48

Tabla 5.14: *Mejores resultados de la clasificación offline con las características extraídas de un método en concreto del conjunto de datos de Training con la partición de datos P3 de "Ia".*

Como se puede apreciar en la Tabla anterior 5.14, el menor y mejor porcentaje de error (13,09 %), es alcanzado con LDA y un vector de características formado por los 11 primeros PCs de todos los 6 canales.

El segundo mejor porcentaje de error (15,48 %), se logra con una SVM con *kernel* Lineal y un vector compuesto por las componentes de baja frecuencia extraídas de los 6 canales, mediante la descomposición Wavelet de un solo nivel, utilizando una onda Wavelet Daubechuies (db4).

Al utilizar los valores del nivel DC y la medida de la Varianza en segmentos de las señales de SCP de los canales 1 y 2, se obtienen resultados satisfactorios con un costo computacional bajo.

Finalmente, se obtienen porcentajes de errores alrededor del 20 %, con un vector formado por los valores de aproximación al valor del área bajo la curva de las señales de todos los 6 canales y un vector con los parámetros AR de los canales 1,2 y 5, estableciendo un orden de 4.

5.2.1.1.7. Mejores resultados de la clasificación *offline* del conjunto de datos de *Training* con P3 de "*Ia*" utilizando súper vectores.

Después de observar y analizar los resultados obtenidos en las clasificaciones con un método en concreto, proponemos crear vectores heterogéneos o súper vectores (tal como lo indicamos en el Capítulo 3, en la visualización y caracterización de los conjuntos de datos), que contengan características extraídas con los mejores métodos y técnicas (ver la Tabla 5.14). Con ello, buscamos mejorar el desempeño en la clasificación.

La clasificación con los súper vectores, se lleva a cabo utilizando LDA y una SVM con dos diferentes tipos de *kernel*: Lineal y RBF. En la Tabla 5.15 puede verse el porcentaje de error obtenido con los súper vectores.

Clasificación offline *"Ia"*				
Súper vector de características	**Training (P3)**			
	LDA	SVM		
		Lineal	RBF	
DC (8 Segmentos) (Canal 1 y 2) - Integral (Todos los 6 canales)	13,94	-	13,02	γ=2,1 C=2
DC y Varianza (5 Segmentos) (Canal 1 y 2) - Integral (Todos los 6 canales)	14,31	-	16,54	γ=14 C=14
PCA (11 PCs) (Todos los 6 canales)	13,09	-	10,84	γ=2,7 C=2
DWT (db4) - PCA (11 PCs) (Todos los 6 canales)	13,02	-	10,79	γ=2,6 C=2
DWT (db4) - PCA (9 PCs)(Todos los 6 canales) - DC (Sin segmentos) (Todos los 6 canales)	14,23	-	10,62	γ=3,8 C=2
DWT (db4) - PCA (9 PCs) (Todos los 6 canales) - DC (Sin segmentos) (Canal 1 y 2)	12,62	-	10,44	γ=2,7 C=2
AR(5) (Canal 1,2 y 5) - DC (8 Segmentos) (Canal 1 y 2) - Integral (Todos los 6 canales)	-	12,18	11,45	γ=4,5 C=2

Tabla 5.15: *Mejores resultados de la clasificación offline con súper vectores con el conjunto de Training (P3) de "Ia".*

Como se puede ver en la Tabla anterior 5.15, los resultados obtenidos en la clasificación usando súper vectores (a excepción de PCA en la línea 3, que es el método en concreto con el que se obtiene un mejor desempeño) es más óptima y eficiente que utilizando vectores que contiene las características extraídas con un método en concreto.

Hemos creado un nuevo súper vector (línea 4), buscando con ello, extraer los PCs de las señales de baja frecuencia (cA) y mejorando levemente el porcentaje de error con respecto a extraer los PCs de las señales brutas (línea 3).

En los últimos tres súper vectores, hemos modificado algunas de sus características (con respecto a la Tabla 5.14), como son, el número de PCs a usar, los valores del nivel DC sin segmentar las señales con (1) canales 1 y 2 y (2) todos los 6 canales y el orden del modelo AR. Dichos cambios, mejoran el desempeño en la clasificación.

La celda sombreada de azul que hace referencia al porcentaje de error 10,44 %, indica que es el menor y mejor valor obtenido con una SVM con *kernel* RBF en el conjunto de entrenamiento (*Training*) con una partición de datos P3. En general con este tipo de clasificador se obtiene un mejor desempeño con los súper vectores. Los valores de los parámetros γ y C de la SVM son establecidos para cada uno de los diferentes súper vectores.

El valor 16,54 %, sombreado de azul, hace referencia al porcentaje de error que se corresponde con el súper vector utilizado en la Tabla 5.16 con el que se obtiene el menor y mejor porcentaje de error entrenando con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*).

Las líneas que se observan (-), indican la ausencia del valor del porcentaje de error obtenido con dicho clasificador, ya que no mejora su desempeño o que simplemente la clasificación no ha sido posible.

5.2.1.2. Resultados de la clasificación *online simulado* del conjunto de datos de *Training* con P3 de *"Ia"* utilizando súper vectores.

Como se menciono al comienzo del Capítulo, los resultados de la clasificación de las señales de SCP en el escenario *online simulado* se representan gráficamente. El propósito de ello, es visualizar en qué tiempo es posible

detectar los eventos (mover un cursor arriba y abajo de la pantalla de un computador). Adquirir un tiempo mínimo y óptimo de detección de los eventos en las señales, es fundamental para el diseño de sistemas BCI fiables que requieren de métodos eficientes en el procesamiento en tiempo real.

Para llevar a cabo la clasificación en el escenario *online simulado*, hemos utilizado los súper vectores de características construidos en el análisis *offline* de la Tabla 5.15. Todas las gráficas que se muestran a continuación, optimizan el tiempo de detección de los eventos ($t < 3,5\ s$).

La aplicación de las diferentes técnicas y métodos de extracción de características en éste tipo de análisis, se hace sobre una porción o trozo de los datos de SCP previamente establecido como se comentó en capítulos anteriores, utilizando una ventana de tiempo de una longitud determinada que se va incrementando en la medida que se desplaza a lo largo de toda la señal.

Con una ventana de tiempo de $t = 0,2\ s$ que equivale a tomar 51 muestras aproximadamente cada vez, es posible observar que en la medida que la ventana de tiempo se va desplazando por la señal y se va incrementando, el porcentaje de error en la clasificación tanto para la "clase 0" como para la "clase 1" va disminuyendo progresivamente hasta llegar a un tiempo t en el que el porcentaje de error tiende a estabilizarse.

5.2.1.2.1. Clasificación *online simulado* con LDA y características obtenidas del nivel DC y la Integral.

En la Figura 5.2 está representada la clasificación en el escenario *online simulado* con LDA y el súper vector de características conformado por los valores del nivel DC obtenidos de 8 segmentos de las señales de los canales 1 y 2 y los valores de aproximación al valor del área bajo la curva (integral) de todos los 6 canales.

Como puede verse, con LDA es posible detectar los dos eventos con un porcentaje de error global de las 2 clases de 13,5 %, en un tiempo de $t = 2,3\ s$, lo que indica un ahorro en tiempo de $t = 1,2\ s$.

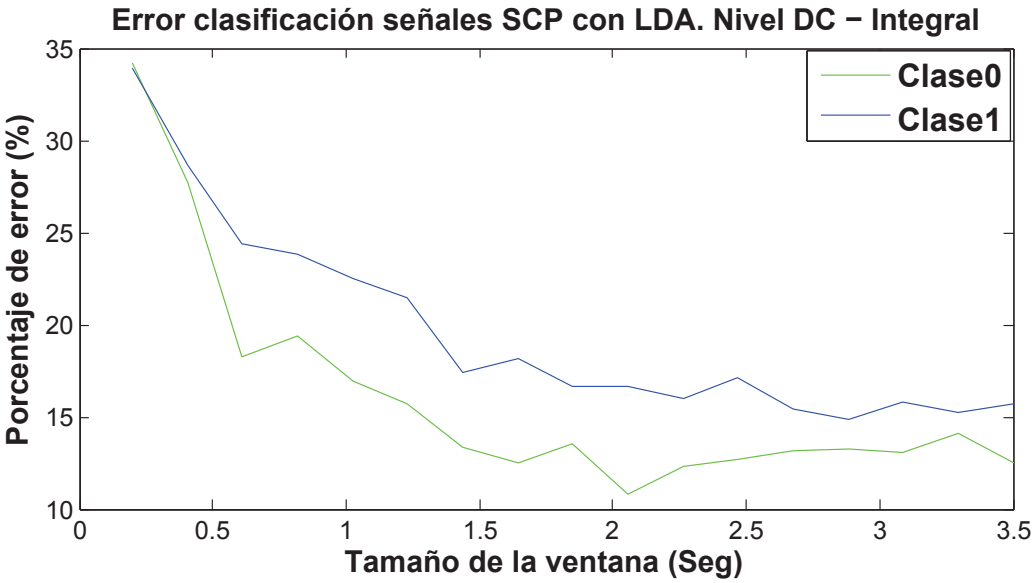

Figura 5.2: *Clasificación online simulado con LDA. Súper vector de características compuesto por el valor DC y el valor de la Integral para el conjunto de datos de Training de "Ia".*

5.2.1.2.2. Clasificación *online simulado* con LDA y características obtenidas de DWT, PCA y DC.

En la Figura 5.3 está representada la clasificación con LDA en el escenario *online simulado* y el súper vector de características conformado por los 9 primeros PCs obtenidos de las componentes de bajas frecuencias (cA) y el valor del nivel DC de las señales en toda su extensión (sin realizar segmentos) de todos los 6 canales.

Como puede verse en la Tabla 5.15, en el escenario *offline* se obtiene un porcentaje de error menor al utilizar una SVM con un *kernel* RBF (10,62 %) que al utilizar LDA (14,23 %).

En el escenario *online simulado* representado en la Figura 5.3, mostramos la clasificación con LDA, ya que es posible detectar los dos eventos con un porcentaje de error global de las 2 clases de 14 %, en un tiempo de $t = 2,9$ s, lo que indica un ahorro en tiempo de $t = 0,6$ s.

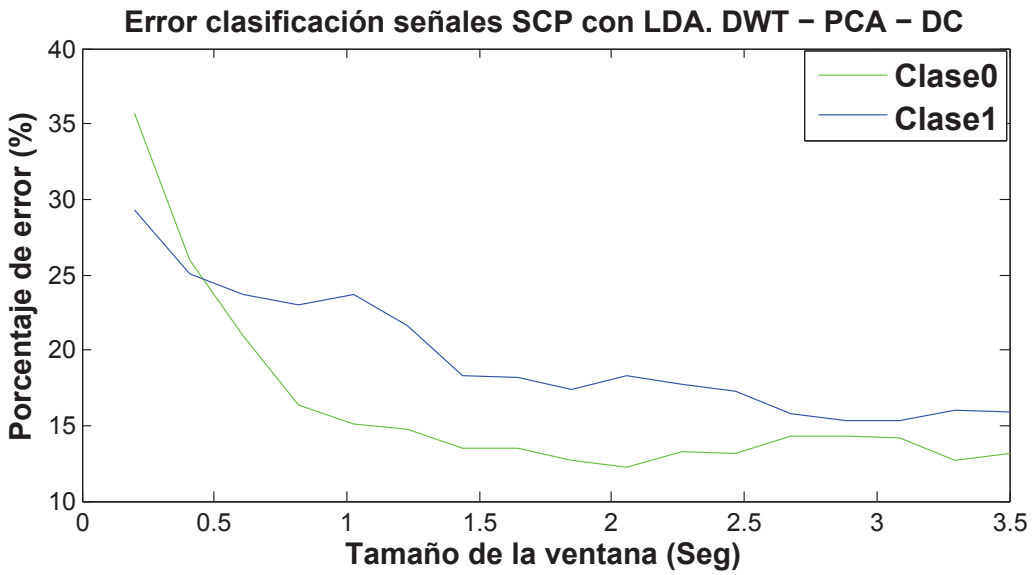

Figura 5.3: *Clasificación online simulado con LDA. Súper vector de características compuesto por los PCs de los cA y el valor DC de todos los 6 canales para el conjunto de datos de Training de "Ia".*

5.2.1.2.3. Clasificación *online simulado* con una SVM Lineal y características obtenidas de AR, DC y la Integral.

En la Figura 5.4 está representada la clasificación con una SVM con *kernel* Lineal en el escenario *online simulado* y el súper vector de características conformado por los parámetros AR de los canales 1, 2 y 5, con un orden establecido en 5, los valores del nivel DC obtenido de 8 segmentos de las señales de los canales 1 y 2 y los valores de aproximación al valor del área bajo la curva (integral) de todos los 6 canales.

Como se puede observar, usando una SVM con *kernel* Lineal, se logra detectar los eventos de la "clase 0" y la "clase 1", con un porcentaje de error global de 11,25 %, en un tiempo de $t = 2,3\ s$, lo que indica un ahorro significativo en tiempo de $t = 1,2\ s$.

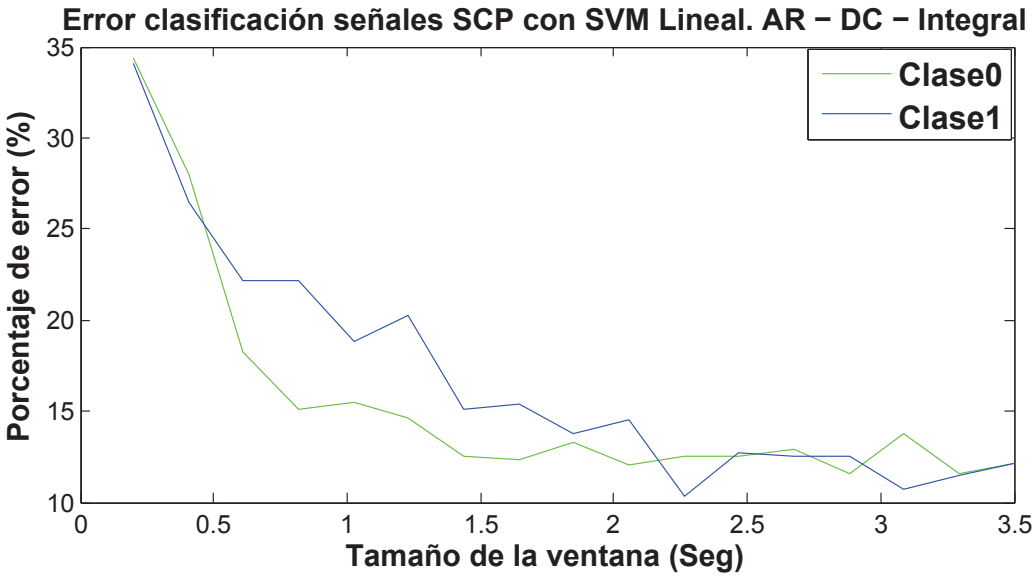

Figura 5.4: *Clasificación online simulado con una SVM Lineal. Súper vector de características compuesto por AR, nivel DC y la Integral para el conjunto de datos de Training de "Ia".*

5.2.1.3. Mejores resultados de la clasificación *offline*, entrenando con el conjunto de *Training* y testeando con el conjunto de *Testing* de *"Ia"* utilizando súper vectores.

A continuación presentamos los resultados de la clasificación en el escenario *offline* del conjunto de datos *"Ia"*, entrenando con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*).

Dicha clasificación, se ha llevado a cabo con súper vectores (de la misma forma que se hizo en el conjunto de entrenamiento (*Training*) con partición de datos P3 para entrenar y validar), tal y como se muestra en Tabla 5.16.

La celda sombreada de azul que hace referencia al porcentaje de error 10,25 %, indica que es el menor y mejor valor obtenido con una SVM con *kernel* RBF. En general con este tipo de clasificador se obtiene un mejor desempeño con los súper vectores. Los valores de los parámetros γ y C de la SVM son establecidos para cada uno de los diferentes súper vectores.

El valor 12,65 %, sombreado de azul, hace referencia al porcentaje de error que se corresponde con el súper vector utilizado en la Tabla 5.15 con el que se obtiene el menor y mejor porcentaje de error con el conjunto de entrena-

miento (Training) con una partición de datos P3 (75 % para entrenar y un 25 % para validar el modelo).

Las líneas que se observan (-), indican la ausencia del valor del porcentaje de error obtenido con dicho clasificador, ya que no mejora su desempeño o que simplemente la clasificación no ha sido posible.

Clasificación offline *"Ia"*				
Súper vector de características	**Training - Testing**			
	LDA	SVM		
		Lineal	RBF	
DC (8 Segmentos)(Canal 1 y 2) - Integral (Todos los 6 canales)	16,37	-	11,24	γ=4,4 C=1
DC y Varianza (5 Segmentos) (Canal 1 y 2) - Integral (Todos los 6 canales)	17,74	-	10,25	γ=16 C=2
PCA (1 PC) (Todos los 6 canales)	11,26	-	11,28	γ=4 C=1
DWT (db4) - PCA (1 PC) (Todos los 6 canales)	11,26	-	11,28	γ=3,9 C=1
DWT (db4) - PCA (Todos los 6 canales) - DC (Sin segmentos) (Todos los 6 canales)	35,05	-	18,08	γ=3,8 C=2
DWT (db4) - PCA (1 PC) (Todos los 6 canales) - DC (Sin segmentos) (Canal 1 y 2)	20,86	-	12,65	γ=2,7 C=2
AR(5) (Canal 1,2 y 5) - DC (8 Segmentos) (Canal 1 y 2) - Integral (Todos los 6 canales)	-	11,25	11,91	γ=4 C=1

Tabla 5.16: *Mejores resultados de la clasificación offline con súper vectores entrenando con el conjunto de Training y testeando con el conjunto de Testing de "Ia".*

5.2.1.4. Resultados de la clasificación *online simulado*, entrenando con el conjunto de *Training* y testeando con el conjunto de *Testing* de *"Ia"* utilizando súper vectores.

La clasificación en el escenario *online simulado*, sigue el mismo procedimiento realizado en el conjunto de datos de entrenamiento *Training* usando la partición de datos P3, descrito anteriormente en el mismo escenario de análi-

sis. Se muestran las gráficas que optimizan la detección de los dos eventos (mover cursor arriba y abajo de una pantalla de computador) en un tiempo $t < 3,5\ s$.

5.2.1.4.1. Clasificación *online simulado* con una SVM con *kernel* RBF y características obtenidas del nivel DC y la Integral.

En la Figura 5.5 se representa la clasificación en el escenario *online simulado* con una SVM con *kernel* RBF, cuyos valores de los parámetros γ y C están establecidos en 4,4 y 1 respectivamente. El súper vector de características está formado por los valores del nivel DC obtenidos de 8 segmentos de las señales de los canales 1 y 2 y los valores de aproximación al valor del área bajo la curva (integral) de todos los 6 canales.

Como se puede observar, usando una SVM con *kernel* RBF, es posible detectar los eventos de la "clase 0" y la "clase 1", con un porcentaje de error global de 11 %, en un tiempo de $t = 2,6\ s$, lo que indica un ahorro en tiempo de $t = 0,9\ s$.

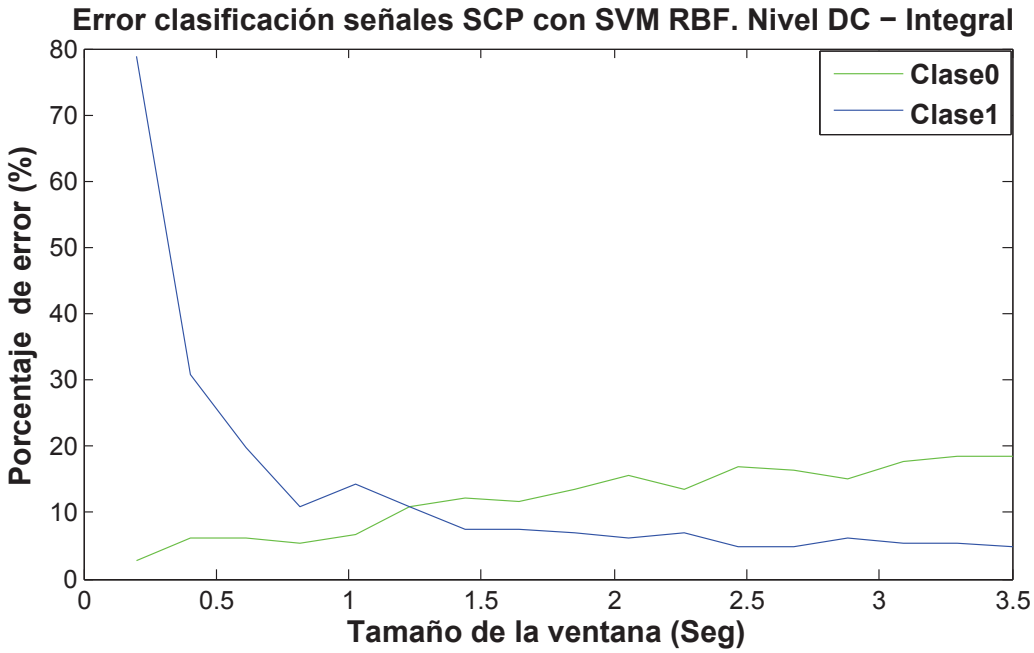

Figura 5.5: *Clasificación online simulado con SVM RBF. Súper vector de características compuesto por el valor DC y el valor de la Integral entrenando con el conjunto de Training y testeando con el conjunto de Testing de "Ia".*

5.2.1.4.2. Clasificación *online simulado* con una SVM con *kernel* RBF y características obtenidas del nivel DC, la Varianza y la Integral

En la Figura 5.6 se muestra la clasificación en el escenario *online simulado* con una SVM con *kernel* RBF, que toma como valores γ=16,4 y C=2.

El súper vector de características está formado por los valores del nivel DC y la medida de la Varianza obtenidos de 5 segmentos de las señales de los canales 1 y 2 y los valores de aproximación al valor del área bajo la curva (integral) de todos los 6 canales.

Como puede verse, con dicho clasificador es posible detectar los dos eventos con un porcentaje de error global de las 2 clases de 9 %, en un tiempo de $t = 2,3\ s$, lo que indica un ahorro considerable en tiempo de $t = 1,2\ s$.

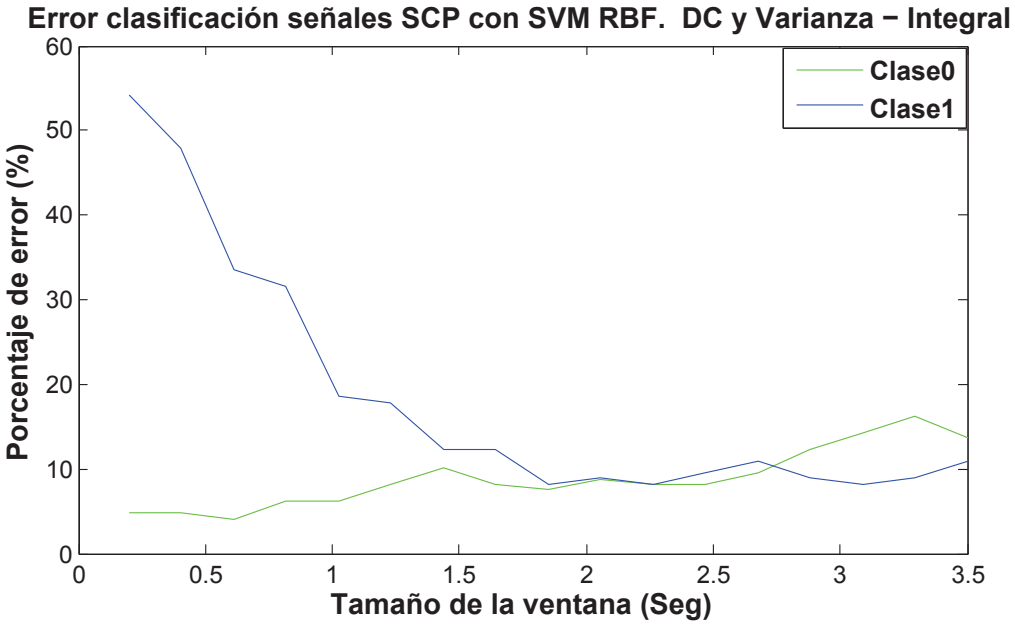

Figura 5.6: *Clasificación online simulado con SVM RBF. Súper vector de características compuesto por el el valor DC, la Varianza y el valor de la Integral entrenando con el conjunto de Training y testeando con el conjunto de Testing de "Ia".*

5.2.1.4.3. Clasificación *online simulado* con LDA y características obtenidas de PCA.

En la Figura 5.7 se representa la clasificación en el escenario *online simulado* con LDA y un vector de características formado solamente por el primer PC usando todos los 6 canales.

Mostramos la clasificación con LDA, ya que es posible detectar los dos eventos con un porcentaje de error global de la "clase 0" y la "clase 1" de 9,1 %, en un tiempo de $t = 2,25\ s$, lo que indica un ahorro en tiempo muy significativo de $t = 1,25\ s$.

Vale destacar que con el súper vector compuesto por características del nivel DC, la Varianza y la Integral (ver Figura 5.6) se obtiene también un porcentaje de error muy similar en un tiempo significativo, con la diferencia del tipo de clasificador utilizado.

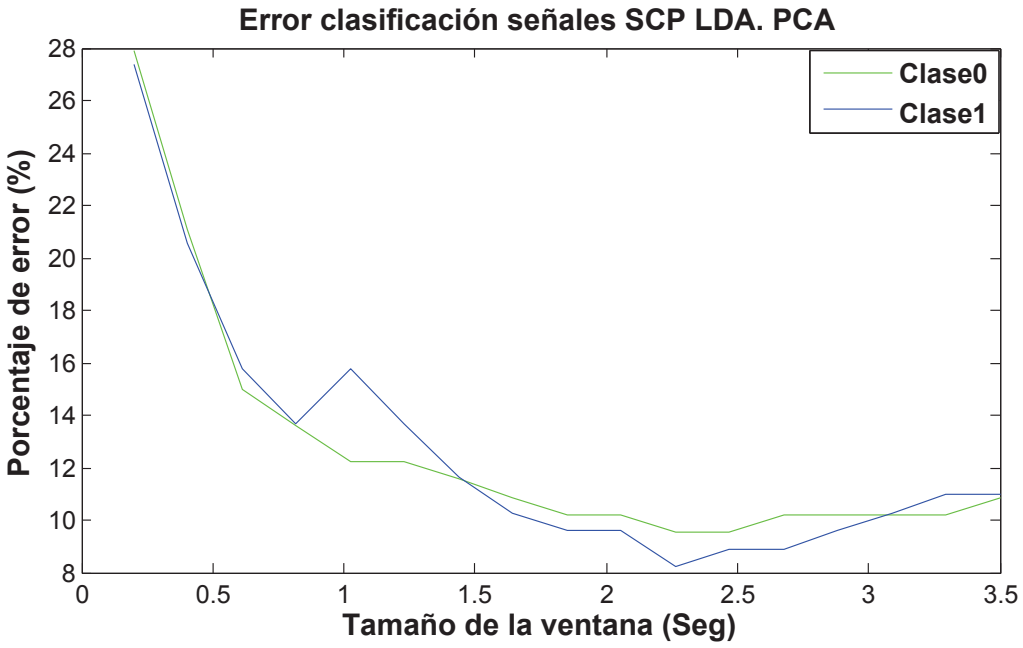

Figura 5.7: *Clasificación online simulado con LDA. Súper vector de características compuesto por los PCs entrenando con el conjunto de Training y testeando con el conjunto de Testing de "Ia".*

5.2.1.5. Resumen general de los mejores resultados en los escenarios *offline* y *online simulado* con los diferentes súper vectores del conjunto de datos "*Ia.*"

En la Tabla 5.17 se puede observar un resumen general de los mejores y menores porcentajes de error obtenidos en la clasificación con los diferentes súper vectores, en los escenarios de análisis *offline* de las Tablas 5.15 y 5.16 y *online simulado* para el conjunto de datos "*Ia*".

Resultados generales clasificación conjunto de datos "*Ia*"				
Escenario	Datos	% Error/ Tiempo	Clasificador	Súper vector
Offline	Training (P3)	10,44 % 3,5 *s*	SVM (RBF)	DWT(db4) - PCA(9 PCs) (Todos los 6 canales) - DC(Sin segmentos) (Canal 1 y 2)
	Testing	10,25 % 3,5 *s*	SVM (RBF)	DC y Varianza(5 Segmentos) (Canal 1 y 2) - Integral (Todos los 6 canales)
Online simulado	Training (P3) Figura 5.4	11,25 % 2,3 *s*	SVM (Lineal)	AR(5) (Canal 1,2 y 5) - DC(8 Segmentos) (Canal 1 y 2) - Integral (Todos los 6 canales)
	Testing Figura 5.6	9 % 2,3 *s*	SVM (RBF)	DC y Varianza(5 segmentos) (Canal 1 y 2) - Integral (Todos los 6 canales)
Ganador de "BCI Competitions II" [73]	Testing	11,30 % 3,5 *s*	LDA	DC(últimos dos tercios cada canal)(canal 1 y 2) - Multitaper de Thomson (canal 4 y 6)

Tabla 5.17: *Resumen general de los mejores resultados en los escenarios offline y online simulado con los diferentes súper vectores del conjunto de datos "Ia' (Tablas 5.15 y 5.16.).*

En la Tabla anterior 5.17, se puede observar, los resultados obtenidos con

el conjunto de entrenamiento (*Training*) y la partición de datos P3 así como los resultados alcanzados entrenando con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*). Mostramos también, el porcentaje de error obtenido por el Ganador de "BCI Competitions II" [2].

Tal y como se puede observar, el desempeño en la clasificación usando súper vectores llevada a cabo en este TFM para el conjunto de datos "*Ia*", mejora entrenando con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*) con respecto al resultado obtenido por el Ganador de "BCI Competitions II" [2].

Obtenemos menores porcentajes de error en la clasificación, entrenando con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*) en ambos escenarios de análisis. Ésta mejora puede ser debido a un aprendizaje por parte del sujeto, ya que los ensayos del conjunto de test (*Testing*) se registraron después (en el segundo día) de las sesiones de grabación de los datos de entrenamiento *Training* o posiblemente sea una característica propia del tipo de señal de control SCP estudiada.

Queremos mencionar también, que en un principio se realizaron pruebas filtrando las primeras muestras en las que aparece un pico (variabilidad de voltaje) en cada uno de los 6 canales tanto para la "clase 0" como para la "clase 1". Dichas pruebas demostraron que la exclusión de los picos no alteraba los resultados de la clasificación, por el contrario, al suprimir muestras se incrementaba el porcentaje de error. Por ello, el procesamiento de las señales se realiza a partir de la señal bruta y completa.

Finalmente, queremos resaltar los resultados obtenidos en el análisis *online simulado*. Entrenando con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*), logramos optimizar el tiempo de detección de los dos eventos (mover un cursor arriba y abajo de la pantalla de un computador) con un porcentaje de error alrededor de 9 %. Con ello, queremos decir, que sólo requerimos $t = 2,3\ s$ del total de la señal proporcionada por "BCI Competition II" que es $t = 3,5\ s$, lo que supone un ahorro significativo de $t = 1,2\ s$.

5.2.2. Resultados del conjunto de datos "*Ib*"

La clasificación con el conjunto de datos "*Ib* ha sido más difícil de llevar a cabo, debido a la alta complejidad de los datos, ya que fueron tomados de una persona con ALS. Sin embargo, seguimos la misma metodología realizada

con el conjunto de datos "*Ia*", presentando los resultados más significativos obtenidos con súper vectores.

Hemos realizado diferentes pruebas con los datos "*Ib*", en las que comprobamos, que el canal 5 (que detectaba los movimientos oculares verticales (vEOG)), no aportaba información significativa que mejorara el desempeño del clasificador. Por ello, hemos omitido dicho canal. De la misma manera, comprobamos que la combinación de los canales 4, 5 y 6, que propusimos en el Capítulo 3, en la visualización y caracterización, tampoco generan resultados aceptables en la clasificación.

5.2.2.1. Resultado de la clasificación *offline* del conjunto de *Training* con P3 de "*Ib*" utilizando súper vectores.

En la Tabla 5.18 se presenta los resultados de los porcentajes de error en la clasificación con el conjunto de entrenamiento (*Training*), realizando una partición de datos P3 (75 % para entrenar y un 25 % para validar el modelo) usando súper vectores. Se han utilizado LDA y una SVM con dos tipos de *kernel*: Lineal y RBF, variando los valores de sus parámetros γ y C.

Como se puede ver, la celda sombreada de azul que hace referencia al porcentaje de error 43,75 %, indica que es el menor valor obtenido con una SVM con *kernel* Lineal en el conjunto de entrenamiento (*Training*) con una partición de datos P3. El súper vector de características está conformado por los parámetros AR de los canales 1,2,3,4,6 y 7, con un orden establecido en 1, los valores del nivel DC obtenido de 8 segmentos de las señales de los canales 1 y 2 y los valores de aproximación al valor del área bajo la curva (integral) de los canales 1,2,3,4,6 y 7.

En general los resultados obtenidos con LDA y una SVM, no son muy diferentes entre si. Los valores de los parámetros γ y C de la SVM son establecidos para cada uno de los diferentes súper vectores.

El valor 49,97 %, también sombreado de azul, hace referencia al porcentaje de error que se corresponde con el súper vector utilizado en la Tabla 5.19 con el que se obtiene el menor porcentaje de error entrenando con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*).

Las líneas que se observan (-), indican la ausencia del valor del porcentaje de error obtenido con dicho clasificador, ya que no mejora su desempeño o que simplemente la clasificación no ha sido posible.

Clasificación offline *"Ib"*				
Súper vector de características	Training (P3)			
	LDA	SVM		
		Lineal	RBF	
DC (8 Segmentos) (Canal 1 y 2) - Integral (Canal 1,2,3,4,6 y 7)	47,07	-	44,34	γ=3 C=2
DC y Varianza (5 Segmentos) (Canal 1 y 2) - Integral (Canal 1,2,3,4,6 y 7)	47,91	-	44,84	γ=2 C=1
PCA (7 PCs) (Canal 1,2,3,4,6 y 7)	48,37	-	49,96	γ=4 C=1
DWT (db4) - PCA (9 PCs) (Canal 1,2,3,4,6 y 7)	48,49	-	49,97	γ=4 C=1
DWT (db4) - PCA (7 PCs) (Canal 1,2,3,4,6 y 7) - DC (Sin segmentos) (Canal 1,2,3,4,6 y 7)	49,90	-	48,54	γ=4 C=1
DWT (db4) - PCA (7 PCs) (Canal 1,2,3,4,6 y 7) - DC (Sin segmentos) (Canal 1 y 2)	48,66	-	49,62	γ=4 C=1
AR(1)(Canal 1,2,3,4,6 y 7) - DC (8 Segmentos)(Canal 1 y 2) - Integral(Canal 1,2,3,4,6 y 7)	-	43,75	45,98	γ=3 C=2

Tabla 5.18: *Mejores resultados de la clasificación offline con súper vectores con el conjunto de Training (P3) de "Ib".*

5.2.2.2. Resultados de la clasificación *online simulado* del conjunto de *Training* con P3 de "*Ib*" utilizando súper vectores.

Los resultados de la clasificación de las señales de SCP en el escenario *online simulado* para el conjunto de datos "*Ib*" se muestran gráficamente de forma similar como se hizo para el conjunto de datos "*Ia*" en dicho escenario.

El objetivo principal en éste análisis, es visualizar en qué tiempo es posible detectar los eventos (mover un cursor arriba y abajo de la pantalla de un computador), utilizando los súper vectores de características definidos en el análisis *offline* de la Tabla 5.18. Se presenta una gráfica en la que es posible visualizar e identificar que el tiempo óptimo de detección de dichos eventos es menor a $t = 4,5\ s$.

Se define una ventana de tiempo de longitud $t = 0,5\ s$ que equivale a tomar

aproximadamente 128 muestras aproximadamente cada vez. Con ésta ventana de tiempo establecida, se puede observar que en la medida que ésta se desplaza por la señal y se va incrementando, el porcentaje de error en la clasificación para la "clase 0" va disminuyendo progresivamente hasta llegar a un tiempo t en el que tiende a estabilizarse. Para la "clase 1", por el contrario, el porcentaje de error aumenta sucesivamente con el tiempo.

5.2.2.2.1. Clasificación *online simulado* con una SVM con *kernel* RBF y características obtenidas del nivel DC y la Integral.

En la Figura 5.8 se representa la clasificación en el escenario *online simulado* con una SVM con *kernel* RBF, que toma como valores γ=3 y C=2. El súper vector de características está conformado por los valores del nivel DC obtenidos de 8 segmentos de las señales de los canales 1 y 2 y los valores de aproximación al valor del área bajo la curva (integral) de los canales 1,2,3,4,6 y 7.

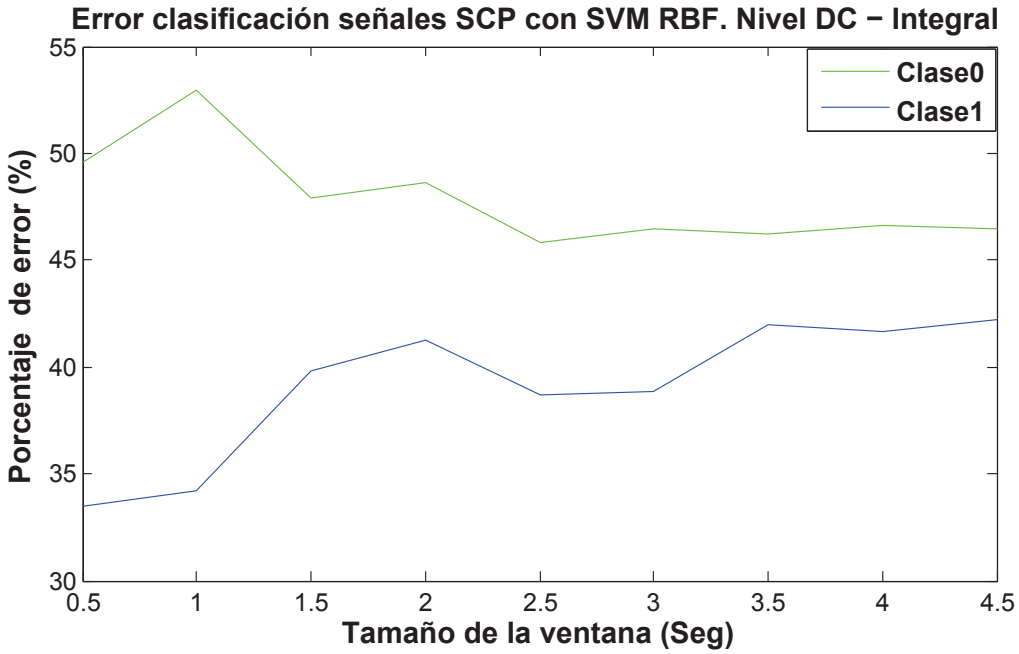

Figura 5.8: *Clasificación online simulado con una SVM con kernel RBF. Súper vector de características compuesto por el valor DC y el valor de la Integral para el conjunto de datos de Training de "Ib"*

Como puede verse en la Figura anterior 5.8, con dicho clasificador es posible detectar los dos eventos con un porcentaje de error global de las 2 clases de 41,9 %, en un tiempo de $t = 2,6\ s$, lo que indica un ahorro muy significativo en tiempo de $t = 1,9\ s$.

5.2.2.3. Resultados de la clasificación *offline* entrenando con el conjunto de *Training* y testeando con el conjunto de *Testing* de "*Ib*" utilizando súper vectores.

A continuación presentamos los resultados de la clasificación en el escenario *offline* del conjunto de datos "*Ib*", entrenando con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*).

Dicha clasificación, se ha llevado a cabo con súper vectores (de la misma forma que se hizo en el conjunto de entrenamiento (*Training*) con partición de datos P3 para entrenar y validar), tal y como se muestra en Tabla 5.19. Hemos modificado los valores de los parámetros de γ y C de la SVM con *kernel* RBF.

Como se puede ver en la Tabla 5.19, la celda sombreada de azul que hace referencia al porcentaje de error 37,78 %, indica que es el menor valor obtenido con una SVM con *kernel* RBF. El súper vector de características está conformado por las componentes de bajas frecuencias (Coeficientes de Aproximación Wavelet) de los canales 1,2,3,4,6 y 7 (tal como lo propusimos en el Capítulo 3, en la visualización y caracterización) y los 3 primeros PCs extraídos de los cA.

El valor 43,89 %, sombreado de azul, hace referencia al porcentaje de error que se corresponde con el súper vector utilizado en la Tabla 5.18 con el que se obtiene el menor porcentaje de error con el conjunto de entrenamiento (Training) con una partición de datos P3 (75 % para entrenar y un 25 % para validar el modelo).

Las líneas que se observan (-), indican la ausencia del valor del porcentaje de error obtenido con dicho clasificador, ya que no mejora su desempeño o que simplemente la clasificación no ha sido posible.

Clasificación offline *"Ib"*				
Súper vector de características	**Training - Testing**			
	LDA	SVM		
		Lineal	RBF	
DC (8 Segmentos) (Canal 1 y 2) - Integral (Canal 1,2,3,4,6 y 7)	45,00	-	42,78	γ=2 C=2
DC y Varianza (5 Segmentos) (Canal 1 y 2) - Integral (Canal 1,2,3,4,6 y 7)	41,11	-	40,55	γ=7 C=1
PCA (6 PCs) (Canal 1,2,3,4,6 y 7)	43,89	-	38,33	γ=4 C=1
DWT (db4) - PCA (3 PCs) (Canal 1,2,3,4,6 y 7)	43,89	-	37,78	γ=4 C=1
DWT (db4) - PCA (11 PCs) (Canal 1,2,3,4,6 y 7) - DC (Sin segmentos) (Canal 1,2,3,4,6 y 7)	43,89	-	40,56	γ=6 C=6
DWT (db4) - PCA (1 PC) (Canal 1,2,3,4,6 y 7) - DC (Sin segmentos) (Canal 1 y 2)	48,89	-	41,11	γ=1 C=2
AR(4) (Canal 1,2,3,4,6 y 7) - DC (8 Segmentos) (Canal 1 y 2) - Integral (Canal 1,2,3,4,6 y 7)	-	43,89	42,78	γ=12 C=2

Tabla 5.19: *Mejores resultados de la clasificación offline con súper vectores entrenando con el conjunto de Training y testeando con el conjunto de Testing de "Ib".*

5.2.2.4. Resultados de la clasificación *online simulado*, entrenando con el conjunto de *Training* y testeando con el conjunto de *Testing* de "*Ib*" utilizando súper vectores.

La clasificación en el escenario *online simulado*, sigue el mismo procedimiento realizado en el conjunto de datos de entrenamiento *Training* usando la partición de datos P3, descrito anteriormente en el mismo escenario de análisis. Se presenta una gráfica en la que es posible visualizar e identificar que el tiempo óptimo de detección de los eventos es menor a $t = 4,5\ s$.

5.2.2.4.1. Clasificación *online simulado* con una SVM con *kernel* RBF y características obtenidas del nivel DC y la Integral.

En la Figura 5.9 se representa la clasificación en el escenario *online simulado* con una SVM con *kernel* RBF, cuyos valores de los parámetros γ y C están establecidos en 2 y 2 respectivamente.

El súper vector de características está formado por los valores del nivel DC obtenidos de 8 segmentos de las señales de los canales 1 y 2 y los valores de aproximación al valor del área bajo la curva (integral) de los canales 1,2,3,4,6 y 7.

Como se puede observar, usando dicho clasificador, es posible detectar los eventos de la "clase 0" y la "clase 1 ", con un porcentaje de error global de 39,95 %, en un tiempo de $t = 4,0\ s$, lo que indica un ahorro en tiempo de $t = 0,5\ s$.

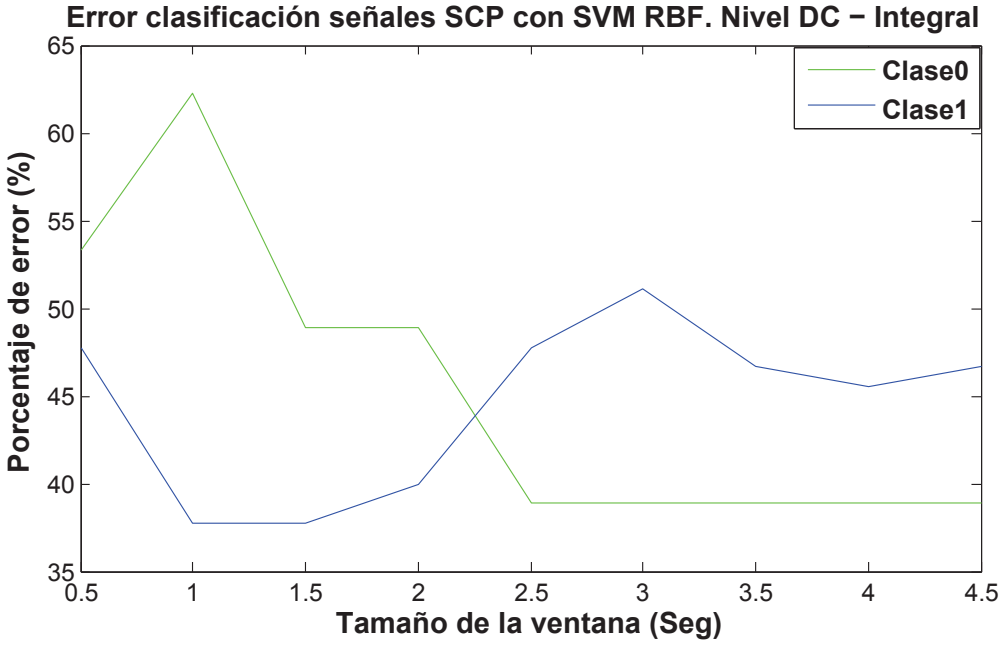

Figura 5.9: *Clasificación online simulado con SVM RBF. Súper vector de características compuesto por el valor DC y el valor de la Integral entrenando con el conjunto de Training y testeando con el conjunto de Testing de "Ib".*

5.2.2.5. Resumen general de los mejores resultados en los escenarios *offline* y *online simulado* con los diferentes súper vectores del conjunto de datos "*Ib*".

En la Tabla 5.20 se puede observar un resumen general de los mejores y menores porcentajes de error obtenidos en la clasificación con los diferentes súper vectores, en los escenarios de análisis *offline* de las Tablas 5.18 y 5.19 y *online simulado* para el conjunto de datos "*Ib*".

Resultados generales clasificación conjunto de datos "*Ib*"				
Escenario	Datos	% Error/ Tiempo	Clasificador	Súper vector
Offline	Training (P3)	43,75 % 4,5 *s*	SVM (Lineal)	AR(1)(Canal 1,2,3,4,6 y 7) - DC(8 Segmentos)(Canal 1 y 2) - Integral(Canal 1,2,3,4,6 y 7)
	Testing	37,78 % 4,5 *s*	SVM (RBF)	DWT(db4) - PCA(3 PCs) (Canal 1,2,3,4,6 y 7)
Online simulado	Training (P3) Figura 5.8	41,90 % 2,6 *s*	SVM (RBF)	DC(8 Segmentos) (Canal 1 y 2) - Integral (Canal 1,2,3,4,6 y 7)
	Testing Figura 5.9	39,95 % 4,0 *s*	SVM (RBF)	DC(8 Segmentos) (Canal 1 y 2) - Integral (Canal 1,2,3,4,6 y 7)
Ganador de "BCI Competitions II" [26]	Testing	45,60 % 4,5 *s*	LDA	DC(últimos dos tercios cada canal)(canal 1 y 2) - Multitaper de Thomson (canal 4 y 6)

Tabla 5.20: *Resumen general de los mejores resultados en los escenarios offline y online simulado con los diferentes súper vectores del conjunto de datos "Ib" (Tablas 5.18 y 5.19)*

Se puede observar en la Tabla anterior 5.20, los resultados obtenidos con el conjunto de entrenamiento (*Training*) y la partición de datos P3 así co-

mo los resultados alcanzados entrenando con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*). Mostramos también, el porcentaje de error obtenido por el Ganador de "BCI Competitions II" [2].

Tal y como se puede ver, el desempeño en nuestra clasificación usando el súper vector conformado por las características extraídas con DWT y PCA, entrenando con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*) es de 37,78 %. Aunque es un resultado malo, debido en gran parte, a la alta complejidad de los datos, ya que son señales tomadas de un sujeto con ALS, si logramos minimizar el porcentaje de error con respecto al obtenido por el Ganador de "BCI Competitions II" [2] que es de 45,60 %.

De la misma manera que en el conjunto de datos "*Ia*", con el conjunto de datos "*Ib*" obtenemos menores porcentajes de error en la clasificación, entrenando con el conjunto de entrenamiento (*Training*) y testeando con el conjunto de test (*Testing*) en ambos escenarios de análisis.

Finalmente, en el análisis *online simulado* del conjunto de datos de "*Ib*", nos podemos dar cuenta, que aunque fue posible identificar y detectar los eventos (mover un cursor arriba y abajo de la pantalla de un computador) en un tiempo inferior a $t = 4{,}5\ s$ que fue el proporcionado por "BCI Competition II", no fue posible minimizar el porcentaje de error en la clasificación.

Parte V

Conclusiones y Trabajo Futuro

5.3. Conclusiones

Este TFM consta de dos partes fundamentales, la primera parte, un estudio y análisis del estado del arte de los sistemas BCI, es decir, conceptos específicos para adentrarse en BCI, técnicas de neuroimagen, señales de control en BCI y técnicas y métodos para extraer las características de las señales cerebrales y su posterior clasificación. Una segunda parte que tiene que ver con el estudio, análisis y la aplicación práctica de dichas técnicas y métodos en el procesamiento de señales reales de EEG relacionadas con SCP.

5.3.1. Conclusiones generales del estado del arte

Los sistemas BCI se basan en la interpretación de las intenciones del usuario monitorizando su actividad cerebral. Existen varias técnicas para registrar la actividad cerebral actualmente, EEG es un método no invasivo de bajo coste, portátil y de fácil uso, por lo que suele ser el método más empleado en estos sistemas. A partir del registro del EEG se pueden obtener distintos tipos de señales adecuadas para controlar un sistema BCI, de las cuales, las más significativas son: VEP, Potenciales Evocados P300, SCP y Ritmos Sensoriomotores μ y β.

Los BCI basados en Ritmos Sensoriomotores y en SCP dependen de la capacidad del usuario para controlar su actividad electrofisiológica, como puede ser la amplitud del EEG en una banda de frecuencia específica sobre un área concreta de la corteza cerebral y requieren de un período de entrenamiento intensivo. Un BCI basado en Ritmos Sensoriomotores se basa en un paradigma de dos o más clases de imágenes motoras o tareas mentales que producen cambios en la amplitud de los ritmos sensoriomotores μ (8-12 Hz) y β (16-24 Hz), registrados sobre la zona somatosensorial y motora de la corteza cerebral. Los SCP son cambios lentos de voltaje generados sobre la corteza cerebral, con una duración variable entre 0.5 s y 10 s. Los SCP negativos, están relacionados con la activación cortical asociada a la realización de movimientos y los SCP positivos, están relacionados con la reducción de la activación cortical. Ambos tipos de señales de control, son especialmente útiles para los usuarios con estados avanzados de ALS o cuyos órganos sensoriales se han visto afectados.

Los sistemas BCI basados en Potenciales Evocados P300 y en VEP dependen de la actividad electrofisiológica evocada por estímulos externos y no necesitan de una etapa intensiva de entrenamiento. El potencial P300 es un pico de amplitud que aparece en el EEG aproximadamente unos 300 ms

después de haberse producido un estímulo auditivo o visual poco frecuente. Los potenciales evocados visuales se detectan en el EEG registrado sobre la zona visual de la corteza cerebral tras haberse aplicado un estímulo visual al usuario. Estos potenciales se hacen estables si la tasa de presentación del estímulo visual está por encima de 6 Hz. Cuando el usuario enfoca su mirada en una imagen que parpadea a una frecuencia determinada, es posible detectar dicha frecuencia analizando el espectro de la señal EEG, ya que aumenta la amplitud del SSVEP en la frecuencia de la imagen parpadeante y en su segundo y tercer armónico. Estos tipos de señales de control pueden alcanzar una alta velocidad de transferencia de información y mucho mayor que los Ritmos Sensoriomotores y SCP.

Los sistemas BCI pueden clasificarse atendiendo a distintas características como son su naturaleza invasiva, si necesitan de estimulación externa, por el tipo de característica usada y, finalmente, si necesitan de un periodo de preparación anterior a la realización o no de la tarea mental. En cuanto a la naturaleza de las causas que provocan los distintos patrones de actividad cerebral que se pueden clasificar en endógenas y exógenas, se podría decir que, todos los BCI necesitan de estimulación externa y por tanto todos tienen cierta naturaleza exógena. Los BCI publicados hasta la fecha usan la técnica del *biofeedback*, necesaria para una comunicación plena. Por tanto sin *biofeedback*, no hay posibilidad de comunicación y sin estimulación externa no existe *biofeedback*.

El procesado de la señal se divide en tres etapas: extracción, selección y clasificación de características. La extracción de características selecciona parámetros que aportan información sobre la intención del usuario. Para ello, se emplean es este trabajo métodos en tiempo, en frecuencia y en tiempo-frecuencia, como la WT, los modelos AR y PCA. El método de selección determina qué características aportan la información que permite discriminar mejor entre dos clases. Por último, el algoritmo de clasificación se encarga de identificar el vector formado por las características seleccionadas con una clase determinada y, a su vez, con un comando de control. Estos algoritmos pueden ser tanto lineales como LDA y no lineales como SVM.

Existen diversas alternativas para la extracción de características, ya sea en el dominio del tiempo, en el dominio de la frecuencia, y/o en el dominio de tiempo-frecuencia. Algunas son más adecuadas que otras en función del tipo de señal de control que se utilice y de las características que se quieran resaltar.

5.3.2. Conclusiones particulares respecto al análisis de datos

Las conclusiones más importantes que se derivan después de realizar el análisis de los datos y de aplicar las distintas técnicas y métodos para extraer y clasificar las características de los conjuntos de datos *"Ia"* y *"Ib"* en los escenarios de análisis *offline* y *online simulado* se pueden resumir en los puntos que se muestran a continuación:

- Al utilizar súper vectores, que contienen características extraídas con las mejores técnicas y métodos de extracción, minimizamos las tasas de error y obtenemos desempeños más eficientes en la clasificación.

- Mediante la técnica propuesta de análisis *online simulado* realizado para los dos conjuntos de datos *"Ia"* y *"Ib"*, fue posible minimizar el tiempo para detectar los eventos (mover un cursor arriba y abajo de la pantalla de un computador), logrando con ello, una mejora significativa en el diseño de sistemas BCI en tiempo real, que requieren mejores tiempos de respuesta (Ver columnas % Error/Tiempo, en las Tablas 5.17 y 5.20).

- Tanto el análisis *offline* como el análisis *online simulado*, son necesarios y complementarios para el diseño de algoritmos efectivos en un entorno real, en el que el usuario recibe *feedback* del sistema BCI.

- Una vez realizado el estudio de las técnicas y métodos en el procesado de las señales de SCP en los conjuntos de datos *"Ia"* y *"Ib"* y se han analizado los resultados obtenidos, hemos comparado con los resultados suministrados por los Ganadores de "BCI Competitions II" [2], comprobando que hemos conseguido mejorar el rendimiento tanto en *"Ia"* como en *"Ib"* (Ver columnas % Error/ Tiempo, en las Tablas 5.17 y 5.20). El súper vector utilizado por la SVM (RBF) en *"Ia"* es DC y Varianza (5 Segmentos) (Canal 1 y 2) - Integral (Todos los 6 canales) que ha generado un error de 10,25 % *versus* 11,30 % que es el error que consiguió el ganador de la competición. El súper vector utilizado por la SVM (RBF) en *"Ib"* es DWT(db4) - PCA(3 PCs) (Canal 1,2,3,4,6 y 7) que generó un error de 37,78 % *versus* el error 45,60 % obtenido por el ganador de la competición.

- La extracción de características usando un súper vector conformado por DC y Varianza (5 segmentos) (Canal 1 y 2) - Integral (Todos los 6 canales) con una SVM (RBF), parece ser un buen método en el análisis *online simulado* en el conjunto de datos *"Ia"*, ya que obtuvimos

una tasa de éxito al rededor de 91 % optimizando el tiempo de respuesta del BCI. En un tiempo de $t = 2,3\ s$ logramos detectar los dos eventos (mover cursor arriba y abajo de una pantalla de computador), consiguiendo un ahorro significativo de $t = 1,2\ s$ del tiempo total proporcionado por "BCI Competition II" que es $t = 3,5\ s$ (Ver Figura 5.6).

Con la misma estrategia de súper vectores, en el conjunto de datos *"Ib"*, obtenemos un éxito en la clasificación alrededor de 60,05 %. Con un súper vector, conformado por DC (8 segmentos) (Canal 1 y 2) - Integral (Todos los 6 canales) con una SVM (RBF), optimizamos el tiempo de respuesta del BCI. Logramos detectar dichos eventos, en un tiempo de $t = 4,0\ s$ del total del tiempo propuesto por "BCI Competition II" que es de $t = 4,5\ s$ (Ver Figura 5.9).

- En el conjunto de datos *"Ia"*, el rendimiento de la clasificación es bastante bueno, teniendo en cuenta que las grabaciones de EEG y las señales del cerebro son demasiado ruidosas y son no estacionarias. El rendimiento de la clasificación en el conjunto de datos *"Ib"* es malo, ya que no hay un patrón evidente para clasificar adecuadamente. Esto nos demuestra que para las personas con ALS, se deben crear metodologías especiales para el procesamiento de datos para su utilización efectiva en BCI.

- Las pruebas en el análisis *offline* consistieron básicamente, en determinar los algoritmos de procesamiento, el tipo de clasificador y la elección de los súper vectores a utilizar en el análisis *online simulado*, para diferenciar satisfactoriamente los eventos (mover cursor arriba o abajo de una pantalla de computador) contando con toda la señal de EEG en su extensión. En las pruebas realizadas en el análisis *online simulado*, los resultados fueron emulados para obtener una estimación de los resultados como si el usuario hubiera utilizado el sistema en tiempo real. En este escenario de análisis, las señales son analizadas y procesadas por segmentos o trozos de un determinado tamaño en el tiempo (ventana). La ventana se va incrementando y desplazando a lo largo de toda la señal de EEG con el objetivo de determinar el segmento mínimo y óptimo que nos permita detectar los eventos con el menor porcentaje de error.

5.4. Líneas futuras

Como mejoras para un proyecto futuro basado en este, podemos proponer:

- Utilizar un modelo Autorregresivo Multivariable Adaptativo para el diseño de sistemas BCI *online* más eficaces que permita obtener los coeficientes AR de las señales no estacionarias.

- Teniendo en cuenta que la DWT es una herramienta con un alto potencial para la extracción de patrones característicos, utilizar dicha herramienta en su forma recursiva, como es la transformada wavelet en paquetes TWP, considerando el nivel de descomposición, el tipo de wavelet madre adecuada para el análisis y, finalmente, qué bandas de frecuencia de las generadas en el proceso de descomposición wavelet realmente determinan o contienen la información necesaria para detectar los eventos de las señales de SCP.

- Estudiar diferentes tipos de algoritmos ICA que permitan descomponer los datos de las señales cerebrales de EEG eliminando los artefactos oculares que están inmersos en dichas señales.

- Hemos manejado una SVM con un *Kernel* RBF, sin embargo hay más posibilidades como utilizar un *Kernel* cuadrático, polinómico o perceptrón multicapa.

- Utilizar otros métodos de clasificación que permitan determinar, a partir de las características seleccionadas, a qué clase de tarea mental pertenece una muestra concreta, tales como, Redes neuronales, K-vecinos más cercanos, Regresión logística ó Modelos de mezclas gaussianas.

- Hacer un análisis más explicito, en la estimación del coste computacional de cada una de las técnicas y métodos de extracción de características estudiados en este TFM.

- Aplicar y demostrar que los algoritmos utilizados en el procesado de las señales de SCP, así como la técnica del análisis *online simulado* podrían ser apropiados para extraer las características en otro tipo de señales de control, tales como, en los Potenciales Evocados P300, en los SSVEP o en los ritmos μ y β.

- Crear una interfaz que fuera completamente adaptativa y que contenga clasificadores adaptativos *online*, con el propósito de suministrar *feedback* en ciclo cerrado a sujetos sanos inexpertos en las primeras etapas de uso. Con ello, podría eliminarse las sesiones iniciales de entrenamiento sin *feedback*, por lo que disminuiría el tiempo total de aprendizaje y los sujetos serían capaces de encontrar una estrategia de funcionamiento directamente con *feedback*.

Parte VI

Apéndices

Apéndice A

Lista de siglas

ALS: Amyotrophic Lateral Sclerosis
ANN: Artificial Neural Network
AR: Autorregresivos
ARMA: AutoRegressive Moving Average Models

BCI: Brain Computer Interface
BSS: Blind Source Separation

CSP: Common Spatial Pattern
CT: Computerized Tomography
CWT: Continuous Wavelet Transform

DC: Direct Current
DET: Detection Error Tradeoff
DFT: Discrete Fourier transform
DWT: Discrete Wavelet Transform

ECG: Electrocardiography
ECoG: Electrocorticogram
EEG: Electroencephalography
EMG: Electromyography
EOG: Electrooculography
EP: Evoked Potential
ERD: Event Related Desynchronization
ERP: Event Related Potentials
ERS: Event Related Synchronization

FFT: Fast Fourier transform
fMRI: functional Magnetic Resonance Imaging
FT: Fourier Transform

GSR: Galvanic Skin Response

ICA: Independent Component Analysis
ITR: Information Transfer Rate

k-NN: Nearest Neighbour

LDA: Linear Discriminant Analysis

MA: Media Móvil
MEG: Magnetoencephalography
MI: Motor Imaginary
MF: Matched Filtering
MLP: Multilayered Perceptron
MRI: Magnetic Resonance Imaging
MTM: Thomson Multi Taper Method

NIRS: Near Infrared Spectroscopy

PC: Principal Component
PCA: Principal Component Analysis
PET: Positron Emission Tomography
PSD: Power Spectral Density

RBF: Radial Basis Function
ROC: Receiver Operator Characteristic

SCP: Slow Cortical Potential
SCI: Spinal Cord Injury
SNR: Signal to Noise Ratio
SPECT: Single Photon Emission Computed Tomography
SSVEP: Steady State Visual Evoked Potential
STFT: Short Time Fourier Transform
SVM: Support Vector Machine

TTD: Thought Translation Device
TFM: Trabajo de Fin de Máster

VEP: Visually Evoked Potential

WT: Wavelet Transform

Bibliografía

[1] *10-20 system (EEG).* http://en.wikipedia.org/wiki/10-20.

[2] *BCI Competition II.* http://www.bbci.de/competition/ii/.

[3] Development of a new rehabilitation system based on a brain-computer interface using near-infrared spectroscopy.

[4] *A diagrammatic representation of the presented BCI system.* http://www.plosone.org/article/info.

[5] *Diferencias en las bases físicas entre EEG y MEG.* http://meg.aalip.jp/vsEEG/vsEEGE.html.

[6] *First mind controlled bionic leg.* http://exoskeletonsc.wordpress.com.

[7] *The future now. With Electrodes Implanted In The Retina, Blind Patient Can Read.* http://www.popsci.com/technology/article/2012-11/electrodes-implanted-retina-blind-patient-can-see-braille-letters.

[8] *La primera mano biónica que siente en tiempo real.* http://www.abc.es/salud/noticias/20140205/abci-hombre-bionico-tiene-mano-201402052126.html.

[9] *Procedimiento para realizar PET.* http://www.cns.nyu.edu/ david/courses/perception/lecturenotes/neuroimaging/neuroimaging.html.

[10] *Red Neuronal Artificial.* http://commons.wikimedia.org/wiki/File:RedNeuronalArtificial.png.

[11] *Resultados de la clasificación: Data sets Ia BCI Competition II.* http://www.bbci.de/competition/ii/results/index.html.

[12] *Resultados de la clasificación: Data sets Ib BCI Competition II.* http://www.bbci.de/competition/ii/results/index.html.

[13] *scikit-learn. Machine Learning in Python.* http://scikit-learn.org/stable/.

[14] *Walk Again Project, address=http://virtualreality.duke.edu/project/walk-again-project/.*

[15] *Neural Engineering.* 2013.

[16] Abrego, P., and Universidad Autónoma Metropolitana (México, D. U. I. *Electrofisiologia Humana: Un Enfoque Para Ingenieros.* Libros de texto, manuales de prácticas y antologías. UAM, Unidad Iztapalapa, 1997.

[17] Akram, F., Han, H.-S., and Kim, T.-S. A p300-based word typing brain computer interface system using a smart dictionary and random forest classifier. In *ICCGI 2013, The Eighth International Multi-Conference on Computing in the Global Information Technology* (2013), pp. 106–109.

[18] Alwasiti, H. H., Aris, I., and Jantan, A. Brain computer interface design and applications: Challenges and future. *World Applied Sciences Journal 11*, 7 (2010), 819–825.

[19] Bayliss, J. D. *A flexible brain-computer interface.* PhD thesis, University of Rochester, 2001.

[20] Bell, A. J., and Sejnowski, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural computation 7*, 6 (1995), 1129–1159.

[21] Bernier, R., Dawson, G., Webb, S., and Murias, M. Eeg mu rhythm and imitation impairments in individuals with autism spectrum disorder. *Brain and cognition 64*, 3 (2007), 228–237.

[22] Birbaumer, N. Slow cortical potentials: Plasticity, operant control, and behavioral effects. *The Neuroscientist 5*, 2 (1999), 74–78.

[23] Birbaumer, N. Breaking the silence: brain–computer interfaces (bci) for communication and motor control. *Psychophysiology 43*, 6 (2006), 517–532.

[24] Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., and Flor, H. A spelling device for the paralysed. *Nature 398*, 6725 (1999), 297–298.

[25] Birbaumer, N., Kubler, A., Ghanayim, N., Hinterberger, T., Perelmouter, J., Kaiser, J., Iversen, I., Kotchoubey, B., Neumann, N., and Flor, H. The thought translation device (ttd) for completely paralyzed patients. *Rehabilitation Engineering, IEEE Transactions on 8*, 2 (2000), 190–193.

[26] Bostanov, V. Bci competition 2003-data sets ib and iib: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *Biomedical Engineering, IEEE Transactions on 51*, 6 (2004), 1057–1061.

[27] Burke, D. P., Kelly, S. P., de Chazal, P., Reilly, R. B., and Finucane, C. A parametric feature extraction and classification strategy for brain-computer interfacing. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 13*, 1 (2005), 12–17.

[28] Cabrera, A. R. Feature extraction and classification for brain-computer interfaces. *Aalborg: Aalborg University, Denmark* (2009).

[29] Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., Patil, P. G., Henriquez, C. S., and Nicolelis, M. A. Learning to control a brain–machine interface for reaching and grasping by primates. *PLoS biology 1*, 2 (2003), e42.

[30] Chen, C., Davis, R. A., and Brockwell, P. J. Order determination for multivariate autoregressive processes using resampling methods. *Journal of multivariate analysis 57*, 2 (1996), 175–190.

[31] Cheng, M., Gao, X., Gao, S., and Xu, D. Design and implementation of a brain-computer interface with high transfer rates. *Biomedical Engineering, IEEE Transactions on 49*, 10 (2002), 1181–1186.

[32] Cohen, D., and Halgren, E. Magnetoencephalography (neuromagnetism). *Encyclopedia of neuroscience* (2003), 1–7.

[33] Cossio, E. G., and Gentiletti, G. G. Interfaz cerebro computadora (icc) basada en el potencial relacionado con eventos p300: análisis del efecto de la dimensión de la matriz de estimulación sobre su desempeño. *Revista Ingenieria Biomedica 2*, 4 (2008), 26–33.

[34] Devore, J. *Probabilidad Y Estadistica Para Ingenieria Y Ciencias/Probability And Statistics For Engineering And Sciences*. Cengage Learning Editores, 2008.

[35] Ding, M., Bressler, S. L., Yang, W., and Liang, H. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biological cybernetics 83*, 1 (2000), 35–45.

[36] Donchin, E., Spencer, K. M., and Wijesinghe, R. The mental prosthesis: assessing the speed of a p300-based brain-computer interface. *Rehabilitation Engineering, IEEE Transactions on 8*, 2 (2000), 174–179.

[37] Donoghue, J. P. Connecting cortex to machines: recent advances in brain interfaces. *nature neuroscience 5* (2002), 1085–1088.

[38] Dornhege, G., Blankertz, B., Curio, G., and Müller, K.-R. Combining features for bci. In *Advances in Neural Information Processing Systems* (2002), pp. 1115–1122.

[39] E, F. Operant conditioning of cortical unit activity. *Science 163*, 3870 (1969), 955–958.

[40] Escolano, C., Murguialday, A. R., Matuz, T., Birbaumer, N., and Minguez, J. A telepresence robotic system operated with a p300-based brain-computer interface: Initial tests with als patients. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE* (2010), IEEE, pp. 4476–4480.

[41] Farina, D., Jensen, W., and Akay, M. *Introduction to Neural Engineering for Motor Rehabilitation.* IEEE Press Series on Biomedical Engineering. Wiley, 2013.

[42] Farwell, L. A., and Donchin, E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology 70*, 6 (1988), 510–523.

[43] Fernandez-Vargas, J., Pfaff, H. U., Rodríguez, F. B., and Varona, P. Assisted closed-loop optimization of ssvep-bci efficiency. *Frontiers in neural circuits 7* (2013).

[44] Fetz, E. E. Volitional control of neural activity: implications for brain–computer interfaces. *The Journal of physiology 579*, 3 (2007), 571–579.

[45] Francis, L. The basics of neural networks demystified. *Contingencies,(11/12 2001)* (2001), 56–61.

[46] Friman, O., Volosyak, I., and Graser, A. Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. *Biomedical Engineering, IEEE Transactions on 54*, 4 (2007), 742–750.

[47] Geddes, L., and Roeder, R. Criteria for the selection of materials for implanted electrodes. *Annals of Biomedical Engineering 31*, 7 (2003), 879–890.

[48] Guardia Serecigni, J., Segura García, L., and Gonzalbo Cirac, B. Aplicaciones de las técnicas de neuroimágen en las conductas adictivas. *Adicciones 12*, 3 (2000).

[49] Gunn, S. R. Support vector machines for classification and regression. *ISIS technical report 14* (1998).

[50] Hammond, D. C. What is neurofeedback? *Journal of Neurotherapy 10*, 4 (2007), 25–36.

[51] He, B. *Neural Engineering*. Bioelectric engineering. Kluwer Academic/Plenum, 2007.

[52] Heinrich, H., Gevensleben, H., and Strehl, U. Annotation: Neurofeedback–train your brain to train behaviour. *Journal of Child Psychology and Psychiatry 48*, 1 (2007), 3–16.

[53] Hinterberger, T., Schmidt, S., Neumann, N., Mellinger, J., Blankertz, B., Curio, G., and Birbaumer, N. Brain-computer communication and slow cortical potentials. *Biomedical Engineering, IEEE Transactions on 51*, 6 (2004), 1011–1018.

[54] Hornero, R., Corralejo, R., and Álvarez, D. Brain-computer interface (bci) aplicado al entrenamiento cognitivo y control domótico para prevenir los efectos del envejecimiento. *Lychnos*, 8 (2012), 29–34.

[55] Hornero, R., Corralejo, R., Álvarez, D., and Martín, L. Diseño, desarrollo y evaluación de un sistema brain computer interface (bci) aplicado al control de dispositivos domóticos para mejorar la calidad de vida de las personas con grave discapacidad. *Trauma 24*, 2 (2013), 117–125.

[56] Hsu, C.-W., Chang, C.-C., Lin, C.-J., et al. A practical guide to support vector classification, 2003.

[57] Huster, R. J., Mokom, Z. N., Enriquez-Geppert, S., and Herrmann, C. S. Brain computer interfaces for eeg neurofeedback: Peculiarities and solutions. *International Journal of Psychophysiology* (2013).

[58] Kelly, S. P., Lalor, E. C., Finucane, C., McDarby, G., and Reilly, R. B. Visual spatial attention control in an independent brain-computer interface. *Biomedical Engineering, IEEE Transactions on 52*, 9 (2005), 1588–1596.

[59] Kennedy, P. R., Kirby, M. T., Moore, M. M., King, B., and Mallory, A. Computer control using human intracortical local field potentials. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 12*, 3 (2004), 339–344.

[60] Klem, G. H., Lüders, H., Jasper, H., and Elger, C. The ten-twenty electrode system of the international federation. the international federation of clinical neurophysiology. *Electroencephalography and clinical neurophysiology. Supplement 52* (1999), 3.

[61] Kübler, A. Brain-computer interfacing: science fiction has come true. *Brain 136*, 6 (2013), 2001–2004.

[62] Kübler, A., Kotchoubey, B., Hinterberger, T., Ghanayim, N., Perelmouter, J., Schauer, M., Fritsch, C., Taub, E., and Birbaumer, N. The thought translation device: a neurophysiological approach to communication in total motor paralysis. *Experimental Brain Research 124*, 2 (1999), 223–232.

[63] Kubler, A., Mushahwar, V., Hochberg, L. R., and Donoghue, J. P. Bci meeting 2005-workshop on clinical issues and applications. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 14*, 2 (2006), 131–134.

[64] Kutz, M. *Standard Handbook of Biomedical Engineering and Design.* McGraw-Hill Handbooks Series. McGraw-Hill Professional Publishing, 2003.

[65] Larsen, E. A. *Classification of EEG Signals in a Brain-Computer Interface System.* PhD thesis, Norwegian University of Science and Technology, 2011.

[66] Lécuyer, A., Lotte, F., Reilly, R. B., Leeb, R., Hirose, M., and Slater, M. Brain-computer interfaces, virtual reality, and videogames. *Computer 41*, 10 (2008), 66–72.

[67] Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G., and Moran, D. W. A brain–computer interface using electrocorticographic signals in humans. *Journal of neural engineering 1*, 2 (2004), 63.

[68] Lin, J., Keogh, E., Wei, L., and Lonardi, S. Experiencing sax: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery 15*, 2 (2007), 107–144.

[69] Malmivuo, J., and Plonsey, R. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields.* Oxford University Press, 1995.

[70] Marks, S., and Dunn, O. J. Discriminant functions when covariance matrices are unequal. *Journal of the American Statistical Association 69*, 346 (1974), 555–559.

[71] Mason, S. G., and Birch, G. E. A general framework for brain-computer interface design. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 11*, 1 (2003), 70–85.

[72] McFarland, D. J., and Wolpaw, J. R. Sensorimotor rhythm-based brain-computer interface (bci): feature selection by regression improves performance. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 13*, 3 (2005), 372–379.

[73] Mensh, B. D., Werfel, J., and Seung, H. S. Bci competition 2003-data set ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals. *Biomedical Engineering, IEEE Transactions on 51*, 6 (2004), 1052–1056.

[74] Minguez, J. Tecnología de interfaz cerebro-computador.

[75] Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J. Wavelet toolbox. *The MathWorks Inc., Natick, MA* (1996).

[76] Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M.-a., Morito, Y., Tanabe, H. C., Sadato, N., and Kamitani, Y. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron 60*, 5 (2008), 915–929.

[77] Moore, M. M. Real-world applications for brain-computer interface technology. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 11*, 2 (2003), 162–165.

[78] Muller, K.-R., Anderson, C. W., and Birch, G. E. Linear and nonlinear methods for brain-computer interfaces. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 11*, 2 (2003), 165–169.

[79] Mussa-Ivaldi, F. A., Casadio, M., and Ranganathan, R. The body-machine interface: a pathway for rehabilitation and assistance in people with movement disorders. *Expert review of medical devices 10*, 2 (2013), 145–147.

[80] Nai-Jen, H., and Palaniappan, R. Classification of mental tasks using fixed and adaptive autoregressive models of eeg signals. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE* (2004), vol. 1, IEEE, pp. 507–510.

[81] Neumann, N., Kaiser, J., Kotchoubey, B., Hinterberger, T., Birbaumer, N. P., et al. Brain-computer communication: self-regulation of slow cortical potentials for verbal communication. *Archives of physical medicine and rehabilitation 82*, 11 (2001), 1533–1539.

[82] Neumann, N., and Kubler, A. Training locked-in patients: A challenge for the use of brain-computer interfaces. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 11*, 2 (2003), 169–172.

[83] Ng, K. B. Visual evoked potential based brain-computer interface.

[84] Nicolas-Alonso, L. F., and Gomez-Gil, J. Brain computer interfaces, a review. *Sensors 12*, 2 (2012), 1211–1279.

[85] Nicolelis, M. A. Actions from thoughts. *Nature 409*, 6818 (2001), 403–407.

[86] Nijboer, F., Furdea, A., Gunst, I., Mellinger, J., McFarland, D. J., Birbaumer, N., and Kübler, A. An auditory brain–computer interface (bci). *Journal of neuroscience methods 167*, 1 (2008), 43–50.

[87] Nowack, W. J. Neocortical dynamics and human eeg rhythms, 1995.

[88] Ono, T., Kimura, A., and Ushiba, J. Daily training with realistic visual feedback improves reproducibility of event-related desynchronisation following hand motor imagery. *Clinical Neurophysiology* (2013).

[89] Palaniappan, R., and Mandic, D. P. Biometrics from brain electrical activity: a machine learning approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 29*, 4 (2007), 738–742.

[90] Palaniappan, R., Syan, C. S., and Paramesran, R. Current practices in electroencephalogram-based brain-computer interfaces.

[91] Pastor, M. C., and BALAÑA, F. J. M. Revisión sobre las técnicas de biofeedback y sus aplicaciones biofeedback techniques and their applications: A review.

[92] Quian Quiroga, R. Single-trial event-related potentials with wavelet denoising: method and applications. In *International Congress Series* (2005), vol. 1278, Elsevier, pp. 429–432.

[93] Rangaswamy, M., Porjesz, B., Chorlian, D. B., Wang, K., Jones, K. A., Bauer, L. O., Rohrbaugh, J., Connor, S. J., Kuperman, S., Reich, T., et al. Beta power in the eeg of alcoholics. *Biological psychiatry 52*, 8 (2002), 831–842.

[94] Rangayyan, R. M. *Biomedical signal analysis.* IEEE press New York, 2002.

[95] Rebsamen, B., Burdet, E., Guan, C., Zhang, H., Teo, C. L., Zeng, Q., Ang, M., and Laugier, C. A brain-controlled wheelchair based on p300 and path guidance. In *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on* (2006), IEEE, pp. 1101–1106.

[96] Sánchez Morillo, D. Procesado y transmisión de señales biomédicas para el diagnóstico de trastornos y enfermedades del sueño.

[97] Schalk, G., Kubanek, J., Miller, K., Anderson, N., Leuthardt, E., Ojemann, J., Limbrick, D., Moran, D., Gerhardt, L., and Wolpaw, J. Decoding two-dimensional movement trajectories using electrocorticographic signals in humans. *Journal of neural engineering 4*, 3 (2007), 264.

[98] Schalk, G., and Mellinger, J. *A Practical Guide to Brain Computer Interfacing with BCI2000: General-Purpose Software for Brain-Computer Interface Research, Data Acquisition, Stimulus Presentation, and Brain Monitoring.* Springer, 2010.

[99] Schneider, F., Rockstroh, B., Heimann, H., Lutzenberger, W., Mattes, R., Elbert, T., Birbaumer, N., and Bartels, M. Self-regulation of slow cortical potentials in psychiatric patients: Schizophrenia. *Biofeedback and self-regulation 17*, 4 (1992), 277–292.

[100] Scholz, M. *Approaches to analyse and interpret biological profile data.* PhD thesis, Universitätsbibliothek, 2006.

[101] Sitaram, R., Caria, A., Veit, R., Gaber, T., Rota, G., Kuebler, A., and Birbaumer, N. Fmri brain-computer interface: a tool for neuroscientific research and treatment. *Computational intelligence and neuroscience 2007* (2007).

[102] Stecklow, M. V., Infantosi, A. F. C., and Cagy, M. Alterações na banda alfa do eletrencefalograma durante imagética motora visual e cinestésica. *Arq Neuropsiquiatr 65*, 4a (2007), 1084–8.

[103] Stone, J. V. *Independent component analysis.* Wiley Online Library, 2004.

[104] Strong, M. J., Grace, G., Orange, J., Leeper, H., Menon, R., and Aere, C. A prospective study of cognitive impairment in als. *Neurology 53*, 8 (1999), 1665–1665.

[105] Thomson, D. J. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE 70*, 9 (1982), 1055–1096.

[106] Ting, W., Guo-zheng, Y., Bang-hua, Y., and Hong, S. Eeg feature extraction based on wavelet packet decomposition for brain computer interface. *Measurement 41*, 6 (2008), 618–625.

[107] Tsoi, A. C., So, D., and Sergejew, A. Classification of electroencephalogram using artificial neural networks. *Advances in neural information processing systems 6* (1994), 1151–1158.

[108] Vapnik, V. *The nature of statistical learning theory.* springer, 2000.

[109] Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S., and Schwartz, A. B. Cortical control of a prosthetic arm for self-feeding. *Nature 453*, 7198 (2008), 1098–1101.

[110] Vinther, M. Independent component analysis of evoked potentials in eeg. *Orsted, DTU. December* (2002).

[111] WANG, X.-Y., JIN, J., ZHANG, Y., and WANG, B. Brain control: Human-computer integration control based on brain-computer interface approach. *Acta Automatica Sinica 39*, 3 (2013), 208–221.

[112] Wang, Y., Wang, R., Gao, X., Hong, B., and Gao, S. A practical vep-based brain-computer interface. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 14*, 2 (2006), 234–240.

[113] Weiskopf, N., Veit, R., Erb, M., Mathiak, K., Grodd, W., Goebel, R., and Birbaumer, N. Physiological self-regulation of regional brain activity using real-time functional magnetic resonance imaging (fmri): methodology and exemplary data. *Neuroimage 19*, 3 (2003), 577–586.

[114] Wolpaw, J. R. Brain–computer interface research comes of age: traditional assumptions meet emerging realities. *Journal of motor behavior 42*, 6 (2010), 351–353.

[115] Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., Donchin, E., Quatrano, L. A., Robinson, C. J., Vaughan, T. M., et al. Brain-computer interface technology: a review of the first international meeting. *IEEE transactions on rehabilitation engineering 8*, 2 (2000), 164–173.

[116] Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. Brain–computer interfaces for communication and control. *Clinical neurophysiology 113*, 6 (2002), 767–791.

[117] Wolpaw, J. R., McFarland, D. J., Vaughan, T. M., and Schalk, G. The wadsworth center brain-computer interface (bci) research and development program. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 11*, 2 (2003), 1–4.

[118] Wood, H. Neural repair and rehabilitation: Achieving complex control of a neuroprosthetic arm. *Nature Reviews Neurology 9*, 2 (2013), 62–62.

[119] Xu, S., Yashchuk, V. V., Donaldson, M. H., Rochester, S. M., Budker, D., and Pines, A. Magnetic resonance imaging

with an optical atomic magnetometer. *Proceedings of the National Academy of Sciences 103*, 34 (2006), 12668–12671.

[120] Yanagisawa, K., Asaka, K., Sawai, H., Tsunashima, H., Nagaoka, T., Tsujii, T., and Sakatani, K. Brain-computer interface using near-infrared spectroscopy for rehabilitation. In *Control Automation and Systems (ICCAS), 2010 International Conference on* (2010), IEEE, pp. 2248–2253.

[121] Zhang, L., He, W., Miao, X., and Yang, J. Dynamic eeg analysis via the variability of band relative intensity ratio: A time-frequency method. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the* (2006), IEEE, pp. 2664–2667.

[122] Zhang, Y., Chen, Y., Bressler, S. L., and Ding, M. Response preparation and inhibition: the role of the cortical sensorimotor beta rhythm. *Neuroscience 156*, 1 (2008), 238–246.